

Herausgeber: Elisabeth Brandenburg, Laura Doria, Alice Gross, Torsten Günzler und Hardy Smieszek

Grundlagen und Anwendungen der Mensch-Maschine-Interaktion

Foundations and Applications of Human-Machine Interaction

10. Berliner Werkstatt Mensch-Maschine-Systeme

10. – 12. Oktober 2013



prometei
graduierntenkolleg

10. Berliner Werkstatt Mensch-Maschine-Systeme
Grundlagen und Anwendungen der Mensch-Maschine-Interaktion
Berlin
10. – 12. Oktober 2013

10th Berlin Workshop Human-Machine Systems
Foundations and Applications of Human-Machine-Interaction
Berlin, Germany
10th – 12th October, 2013

Editors: Elisabeth Brandenburg
Laura Doria
Alice Gross
Torsten Günzler
Hardy Smieszek

Universitätsverlag der TU Berlin

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data are available in the Internet at <http://dnb.dnb.de/>

Universitätsverlag der TU Berlin 2013

<http://www.univerlag.tu-berlin.de>

Fasanenstr. 88, 10623 Berlin

Tel.: +49 (0)30 314 76131 / Fax: -76133

E-Mail: publikationen@ub.tu-berlin.de

License: All contents of this publishing are licensed
under the following Creative-Commons-License agreement
<http://creativecommons.org/licenses/by-sa/3.0/de/>

ISBN: 978-3-7983-2626-2 (Online)

Published online on the Digital Repository of the Technische Universität Berlin:

URL <http://opus4.kobv.de/opus4-tuberlin/frontdoor/index/index/docId/4289>

URN <urn:nbn:de:kobv:83-opus4-42897>

<http://nbn-resolving.de/urn:nbn:de:kobv:83-opus4-42897>

Der vorliegende Tagungsband zur 10. Berliner Werkstatt Mensch-Maschine-Systeme gibt einen Einblick in die aktuelle Forschung im Bereich der Mensch-Maschine-Interaktion. Einen besonderen Fokus stellt das Wechselspiel von Grundlagenforschung und anwendungsbezogener Forschung dar, was sich im breiten Themenspektrum widerspiegelt, welches von theoretischen und methodischen Betrachtungen bis hin zu anwendungsnahen Fragestellungen reicht. Dabei finden Inhalte aus allen Phasen des Forschungsprozesses Beachtung, sodass auch im Rahmen der 10. Berliner Werkstatt MMS wieder sowohl neue Untersuchungskonzepte als auch abschließende Befunde diskutiert werden. Zentrale Themengebiete sind u. a. Fahrer-Fahrzeug-Interaktion, Assistenzsysteme, User Experience, Usability, Ubiquitous Computing, Mixed & Virtual Reality, Robotics & Automation, Wahrnehmungsspezifika sowie Psychophysiologie und Beanspruchung in der Mensch-Maschine-Interaktion.

The proceedings of the 10th Berlin Workshop Human-Machine-Systems provide an insight into the current research in the field of human-machine-interaction. The main focus lies on the interplay between basic and applied research, which is reflected in the wide range of subjects: from theoretical and methodological issues to application oriented considerations. Again all stages of the research process are represented in the contributions of the 10th Berlin Workshop HMS. This means new research concepts as well as final results are subject of this volume. Central topics include driver-vehicle-interaction, assistance systems, user experience, usability, ubiquitous computing, mixed and virtual reality, robotics & automation, perception specifics, as well as psychophysiology and workload in human-machine-interaction.

Keywords: Mensch-Maschine-Systeme, Unterstützungs- und Assistenzsysteme, User Experience, Mixed & Virtual Reality, Usability, Robotics & Automation

Vorwort

Die Berliner Werkstatt Mensch-Maschine-Systeme findet in diesem Jahr zum zehnten Mal statt. Seit den Anfängen der Tagung vor 18 Jahren bis zur Gegenwart hat sich das Konzept eines angeregten fachlichen Dialogs zu wissenschaftlich-technischen Trends mit speziellem Fokus auf die Interaktion zwischen Mensch und Maschine bewährt. Der Name der Tagung weist auf den beabsichtigten Werkstattcharakter der Veranstaltung hin, indem zum einen durch eine angemessene Diskussionszeit nach den Vorträgen die geeigneten Rahmenbedingungen für einen intensiven Gedankenaustausch geschaffen werden. Zum anderen vereint die enge Vernetzung von Wissenschaft und Industrie, von Theorie und Praxis ganz unterschiedliche Sichtweisen auf die aktuellen Forschungsthemen und macht so den Erfolg der Werkstatt aus.

Die 10. Berliner Werkstatt Mensch-Maschine-Systeme fokussiert folgerichtig auf das Wechselspiel von Grundlagenforschung und anwendungsbezogener Forschung im Bereich der Mensch-Maschine-Interaktion und allgemeiner der Mensch-Maschine-Systemforschung. Im Rahmen anwendungsbezogener Forschung werden bisweilen Defizite offenbar, die nur durch Grundlagenforschung zu schließen sind. Umgekehrt ergeben sich aus der Grundlagenforschung bisweilen neue Möglichkeiten für neue Formen der Mensch-Maschine-Interaktion, die dann mittels anwendungsnaher Forschung weiterentwickelt werden.

Im Programm finden sich dementsprechend Werkstattgespräche, die die Grundlagen der Mensch-Maschine-Systemforschung im Bereich der Psychophysiologie und von Wahrnehmungsspezifika aufgreifen. Und im Bereich der anwendungsbezogenen Forschung sind ein besonderer Schwerpunkt die inzwischen großflächig beim Endanwender angekommenen Fahrerassistenzsysteme. Die Tagung bietet dazu eine durchgehende Veranstaltungslinie an, die Vorträge und Werkstattgespräche speziell zu (fehlenden) Möglichkeiten, aber auch Risiken und Auswirkungen von unterstützenden Assistenten während der Fahrt im Fahrzeug, Flugzeug oder Zug bietet. Weitere Werkstattgespräche greifen die Robotics & Automation, Mixed & Virtual Reality, Assistenzsysteme allgemein sowie verschiedene Aspekten der Usability auf. Besonders hinweisen möchten wir auf die beiden Special Sessions: 1) Das europäische Forschungsnetzwerk HAMASYT (Human-Machine Systems in Transportation and Industry) präsentiert seine Aktivitäten mit acht Beiträgen und 2) Kay Behrenbruch und Ludger Schmidt laden zur Vorstellung und Diskussion unter dem Titel „Ubiquitous Computing“ ein.

Ein besonderes Highlight der Tagung ist der eingeladene Vortrag von Prof. Charles Spence, Professor of Experimental Psychology, University of Oxford, Nobelpreisträger für Ernährung (2008), für den Nachweis, dass Essen besser schmeckt, wenn es ansprechend benannt wird. Er beschäftigt sich vor allem mit neuronalen Vorgängen wie der Wirkung von Aufmerksamkeit in der multimodalen Wahrnehmung und wie diese genutzt werden kann in der Konsumpsychologie. Sein Vortrag trägt den Titel "Neuroscience-inspired multisensory design in the technology sector work". Er wird hier seine aktuelle Forschung im Bereich Neuroergonomie vorstellen.

Wie schon in den letzten Jahren präsentieren sich die Poster in entspannter Atmosphäre am Abend bei Wein und Brezeln.

Besonders hinweisen möchten wir noch auf die Aussteller, die sich und ihre Produkte nicht nur in der Ausstellung, sondern auch in einem der parallelen Blöcke präsentieren werden:

- Ergoneers - www.ergoneers.com
- SensoMotoric Instruments GmbH - www.smivision.com
- Smart Eye AB - www.smarteye.se

Darüber hinaus danken wir der Firma Brain Products (www.brainproducts.com) für eine großzügige Spende, mit der sie diese Tagung unterstützt hat.

Die im Programm aufgenommenen hochwertigen Beiträge aus unterschiedlichen Perspektiven lassen nun inzwischen zum zehnten Mal einen regen Austausch an Ergebnissen, Methoden und Ideen zwischen Forschung und Industrie erwarten. Wir freuen uns auf spannende Diskussionen, insbesondere unter der Perspektive des Wechselspiels von Grundlagenforschung und anwendungsbezogener Forschung im Bereich der Mensch-Maschine-Interaktion und allgemeiner der Mensch-Maschine-Systemforschung.

Berlin, Oktober 2013

Prof. Dr. Klaus Gramann

Dr.-Ing. Katja Karrer-Gauß

Prof. Dr.-Ing. Matthias Rötting

Prof. Dr.-Ing. Nele Rußwinkel

Tag 1**Se 1-1 Fahrer-Fahrzeug-Interaktion**

M. Kolbig, S. Müller <i>Mode awareness im Fahrkontext: Eine theoretische Betrachtung</i>	1
M. Domhardt, E. Tunca, I. Zoller, P. Lotz, L. Schmidt <i>Evaluation eines haptischen Touchpads für die Fahrer-Fahrzeug-Interaktion</i>	9
M. Volk, I. Totzke <i>Nutzerakzeptanz von Car-2-X-Technologien: Welche Anforderungen machen Fahrerassistenzsysteme beliebter als andere?</i>	19

Se 1-2 User Experience

M. Minge, L. Riedel, M. Thüring <i>Modulare Evaluation interaktiver Technik. Entwicklung und Validierung des meCUE Fragebogens zur Messung der User Experience.</i>	28
M. Körber, K. Bengler <i>Zur Validität eines User Experience-Fragebogens zur Messung der Bedürfniserfüllung in Produktinteraktionen</i>	37
M. Auricht, E. Dittrich, S. Brandenburg, R. Stark <i>Engineering meet UX: Ingenieurstechnische Absicherungsmethoden treffen auf User Experience Forschung</i>	44

Se 1-3 HAMASYT Special Session Teil I

P. Polet, M. Rötting, C. Jonker, P. Wieringa, F. Vanderhaegen <i>Non-intrusive systems for human state diagnosis based on facial recognition and eye-tracking systems</i>	51
S. Ruff, M. Rötting <i>Impact of increasing workload on facial temperature in a simulated driving task</i>	58
S. Debernard, P. Marangé, F. Gellot, M.-P. Pacaux-Lemoine, T. Poulain, A. Philippot, B. Riera, J.-F. Pétrin <i>A support tool for assisting human diagnoses of command errors detected by a robust filter</i>	66

Se 2-1 Fahrerbeanspruchung während und nach der Fahrt

O. Mantzke, R. Vilimek, C.-C. Carbon <i>Driver Distraction: Weiterentwicklung einer Methode zur Messung von kognitiver Beanspruchung durch Anzeige-und Bedienkonzepte im Fahrzeug</i>	75
M. Helmbrecht, K. Bengler, R. Vilimek <i>Strategies for Efficient Driving with Electric Vehicles</i>	83
K. Karrer-Gauß, M. Issing <i>Effects of telemedical bystander support on helping behaviour after traffic accidents</i>	91

Se 2-2 Ubiquitous Computing Special Session

L. Schmidt, K. Behrenbruch <i>Interdisziplinäre Systemevaluation von Ubiquitous Computing</i>	102
A. Roßnagel, M. Schuldt <i>Die Simulationsstudie als Evaluationsmethode sozial-verträglicher Technikgestaltung</i>	108
K. Behrenbruch, L. Schmidt <i>Gebrauchstauglichkeit im Labor objektiv evaluieren</i>	117
M. Söllner, K. Behrenbruch, H. Hoffmann, J. M. Leimeister <i>Vertrauenswürdige Gestaltung von ubiquitären Systemen – Potentiale des Einsatzes von Methoden des NeuroIS</i>	126
O. Kieselmann, N. Kopal, A. Wacker <i>Interdisziplinäre Sicherheitsanalyse</i>	135

Se 2-3 HAMASYT Special Session Teil II

P. Simon, J. Floris, J.-C. Popieul <i>The SHERPA driving simulator as a technical platform for driver's state monitoring studies</i>	144
N. Rachedi, D. Berdjag, F. Vanderhaegen <i>Probabilistic techniques to diagnose human operator state</i>	152
C. Kolski, Y. Lebrun, B. Rajaonah, S. Lepreux, R. Mandiau <i>Interactive tabletops as new supports for collective diagnosis?</i>	166
F. Vanderhaegen <i>Toward the reverse comic strip concept to diagnose human state based on sound and facial recognition systems</i>	172

Posterbeiträge Werkstatt-Track

N. Backhaus, S. Brandenburg <i>Wie viel ist zu viel? Richtlinien für die Anzahl von Bedienelementen auf kleinen Displays</i>	180
S. Buyer, C. Wittenberg <i>AR und Webpad – nur Spiele(reien)?</i>	186
N. Fricke, K. Oeltze <i>Assistenz für Elektromobilität - Erfassung der Nutzerakzeptanz</i>	191
X. G. Gong, K.-P. Engelbrecht <i>The influence of user characteristics on the quality of judgment prediction models for tablet applications</i>	197
M. Halbrügge <i>ACT-CV: Bridging the Gap between Cognitive Models and the Outer World</i>	204
A. Johnsen, D. Löffler, M. Rötting <i>Mental Models und Image Schemata: Gestaltung komplexer Software mit Hilfe von Mentalen Modellen der Nutzer</i>	211
W. Liu, R. Lorenz, N. Adolphs, B. Hoffmeister, E. Keeve <i>Domain-Map Method for Communication Interface Optimization in Computer Assisted Orthognathic Surgery – a Pioneer Approach</i>	217
A. Maus, M. Junge <i>Vertrauen in Fahrerassistenzsysteme – Gestaltung von Fahrerassistenz und Vertrauenskalibrierung</i>	222
S. Müller, D. Manzey, A. Bleyer, K. Schreiter, A. Voigt, R. Luckner <i>Untersuchung der mentalen Repräsentation von Energiemanagement bei der Flugzeugführung zur Entwicklung eines Pilotenassistenzsystems</i>	230
R. Reckin, S. Brandenburg <i>Discount Usability-Maßnahmen als erste Schritte von KMU auf dem Weg zum Agile Usability Engineering</i>	238
U. Schmuntzsch, S.-C. Freytag <i>HLB und Web 2.0 – Anforderungen und Konzept für eine Expertenplattform</i>	245
E.-M. Skottke, J. Warkentin, T. Krott <i>Cross-cultural Usability: Eine Untersuchung wie Usability von deutschen vs. russischen Nutzern verstanden wird</i>	252
L. Weber, I. Wetzel <i>Eignung unterschiedlicher Versuchsumgebungen für die Erfassung der wahrgenommenen Qualität von Flugzeugkabinen</i>	259

Posterbeiträge Prometei - UCIT

M. Balaud, D. Manzey <i>Kontrasteffekte bei der Benutzung von Likelihood Alarmsystemen</i>	264
E. Dittrich <i>Annotationen für CAD-Modelle in virtueller Umgebung</i>	272
L. Doria, S. Dannehl, M. Kraft <i>Aktivitätsmessung in der Orthesenversorgung der unteren Extremitäten</i>	277
L. Ferreira dos Santos, H. Schmidt, C. Dohle, J. Krüger <i>Avatare in der virtuellen Neurorehabilitation der unteren und oberen Extremitäten</i>	283
A. M. Gross <i>Exploring Surprise: Influencing User Experience through Unexpected Events</i>	289
T. Günzler, D. Manzey <i>Asymmetrische Unsicherheitsakzeptanz im Umgang mit Entscheidungsunterstützungssystemen</i>	294
E. Ivanova, H. Schmidt, J. Krüger <i>Adaptive Regelungsalgorithmen für das robotergestützte motorische Training der oberen Extremitäten für Patienten nach Schlaganfall: Literatur Review</i>	299
M. Luz, D. Manzey <i>Einfluss von Alarmen chirurgischer Assistenzsysteme auf Leistung, Beanspruchung und Situationsbewusstsein von Chirurgen</i>	309
D. Reich <i>Der Einfluss von Immersivität in virtuellen Umgebungen auf Situation Awareness im Fahrzeug</i>	316
R. Reichmuth, R. Schleicher, S. Möller <i>Klassifikation und Detektion des Nutzungskontextes bei Smartphones</i>	322
T. Salmimaa <i>Collaboration Enable Techniques for Complex Enterprise Resource Planning Projects</i>	327
H. Smieszek, N. Russwinkel <i>Micro-cognition and macro-cognition – Trying to bridge the gap</i>	336
O. Tuisku, V. Surakka, V. Rantanen, J. Lekkala <i>Face Interface: Using voluntary gaze direction and facial muscle activations for human-computer interaction</i>	343
H. Venesvirta, V. Surakka, J. Lylykangas, V. Rantanen, O. Špakov, J. Verho, A. Vetek, J. Lekkala <i>Face-hugging device for technology mediated human-human interaction</i>	349
M. Vogel, J. Silvennoinen <i>Visual Design: The Effect of Mere-Exposure in Different UX Phases</i>	356
S. Werk <i>Einsatz von Schlüsselindikatoren für die Berücksichtigung unsicherer Randbedingungen in der Anlagenführung</i>	364
M. Zickerow, M. Kraft <i>Analyse des Nutzungsverhaltens von Chirurgen beim Einsatz von multifunktionalen minimalinvasiven Chirurgieinstrumenten mit Handschaltern und Pedalen</i>	371

Tag 2**Se 3-1 Warnassistent im Fahrzeug - Bedarf und Wirkung**

- G. Weller, F. Heyne, T. Feige, H. Bretschneider, H. Oeser, B. Schlag 376
Die Wirkung gerichteter Warnungen von Fahrerassistenzsystemen auf die Blickzuwendungs- und Reaktionszeiten von Autofahrern
- F. Heyne, D. Deutsch, G. Weller, B. Schlag 383
Auswirkungen warnender Fahrerassistenzsysteme auf das Blickverhalten und die Bremsreaktion abgelenkter Fahrer in Gefahrensituationen
- D. Raudszus, J. Josten, A. Zlocki, L. Eckstein 391
Modellierung der Bremsreaktion des Fahrers zur Bewertung von aktiven Fußgängerschutzsystemen - Bestimmung des Bremsverhaltens im Fahrsimulator
- F. Naujoks, H. Grattenthaler, A. Neukum 401
Fahrerseitiger Unterstützungsbedarf in drohenden Verkehrskonfliktszenarien und Wirksamkeitsuntersuchung frühzeitiger Fahrerinformationen basierend auf kooperativer Umfelderkennung

Se 3-2 Wahrnehmungsspezifika

- J. Mühlstedt, P. Roßner, A. C. Bullinger 412
Die dunkle Seite des Lichts - Diskomfort durch Flicker bei (LED-) Lichtern im Straßenverkehr in Bezug zu peripheren Flimmerverschmelzungsfrequenzen
- E. Dittrich, A. Ostheimer, A. Vick 421
Visuelle Informationsaufnahme im virtuellen 3D-Raum
- A. Hillebrand 427
ATTENDO: Methode zur Erfassung visueller Aufmerksamkeitsverteilung
- S. Arndt, J. Bürglen, J.-N. Antons, R. Schleicher, S. Möller 435
Einfluss der Audiomodalität auf die Wahrnehmung und Qualitätsbeurteilung (audio-)visueller Stimuli

Se 3-3 Robotics & Automation

- V. Nitsch 442
Situation Awareness in Autonomous Service Robots
- S. Kuz, M. P. Mayer, C. M. Schlick 450
Anthropomorphismus in der Gestaltung der Mensch-Roboter-Interaktion in industriellen Umgebungen
- C. von Bernstorff, J. Nachtwei 456
Kooperative Automation im Labor - Vergleich von manueller und teilautomatisierter Prozessüberwachung und -führung
- S. Meier, E. Altendorf, M. Baltzer, F. Flemisch 466
Partizipative Interaktions- und Automationsgestaltung teil- bis hochautomatisierter Fahrzeuge: Ausgewählte Ergebnisse explorativer Nutzerstudien zu H-Mode 1.1

Se 4-2 Usability

K. Sachse, M. Thüring, J. Grauert 474
Usability und Sicherheit – Ein Widerspruch oder zwei Seiten einer Medaille?

S. Pfeffer, T. Maier 483
Validierung der Gebrauchstauglichkeit medizintechnischer Produkte mit Hilfe der summativen Methode iFlow

Se 4-3 Assistenz im Flugverkehr

P. Manske, M. von Schlippe, M. Lange, C. Möhlenbrink 491
Expertenbefragung: Erfassung der situationsbezogenen Wichtigkeit von Informationsquellen am Towerlotsenarbeitsplatz

P. Manske, H. Smieszek, A. Hasselberg, C. Möhlenbrink 498
Entwicklung eines generischen Flughafen-Modells für die effizientere Makrokognitive Modellierung des Mensch-Maschine-Systems der Flughafenverkehrskontrolle mit farbigen Petri-Netzen

H. Smieszek, F. Joeres 506
Prospective decision making in a macro-cognitive model of airport traffic control system (MATriCS) based on coloured petri nets

F. Fortmann, B. Wortelen 514
Supervising the Supervisor: An Approach to Improve Scanning Behaviour during UAV Supervisory Control

Se 5-1 Assistenz im Bahnverkehr

J. Brünger, P. Grossenbacher, P. Zwahlen, J. Zimmermann, K. Gärtner, T. Wäfler, K. Fischer 524
Mensch-Maschine-Interaktion im Betrieb der SBB

Se 5-2 Psychophysiologie

T. Radüntz, G. Freude 531
Erfassung psychomentaler Beanspruchung bei Arbeit mit unterschiedlichen kognitiven Anforderungen mit Hilfe des Elektroenzephalogramms (EEG)

A. Venjakob, J. Bock, F. Hoppe, P. Brandstädter, S. Ruff 539
Eye Tick - Blickinteraktion im öffentlichen Raum

Se 5-3 English Session Interaction Analysis

- P. E. Nurkka, S. Jumisko-Pyykkö 546
A Light-Weight Mixed Method Tool for Remote UX Evaluation: A Case Study of Premium Sports Watch Users
- H. Väättäjä, T. Oja 556
A mixed methods approach to a longitudinal study of user experience in the metals and engineering industry
- J. Schindler, J. Kelsch, M. Heesen, M. Dziennus, G. Temme, M. Baumann 562
A Collaborative Approach for the Preparation of Cooperative Multi-User Driving Scenarios

Tag 3**Se 6-1 Assistenzsysteme**

- A. Schwarze, M. Horn, M. Haiduk, F. Eggert 570
Entwicklung eines Inhome Displays zur Überwachung und Steuerung des Stromverbrauchs
- F. Maugg, S. Schöning 577
MERLiN — A novel multi-touch based assistance system for biotechnological laboratories
- A. Bremer, G. Rinkenauer 583
Advanced tactile feedback in the context of sensorimotor assistance systems: A tapping task study

Se 6-2 Entwickler & Operateur/ Produkt Lifecycle Management

- R. Tille, M. Burmester, K. Schippert 590
Role-Based-Client Workspace – Entwicklung von Dashboard-Interfaces im Product Lifecycle Management (PLM)
- D. Ley, A. Nielsen 600
Approach to support the modeling of workflows by means of unmoderated data collection sessions

Se 6-3 Mixed & Virtual Reality

- C. Wienrich, J. H. Israel, G. Müller-Plath 608
Virtuelle Realitäten in der psychotherapeutischen Behandlung
- M. Wille, B. Grauel, L. Adolph 617
Beanspruchung bei Head-Mounted Displays im Langzeiteinsatz
- J. Paschke, H. Wandke 626
Verringert Assistenz kognitive Fähigkeiten? - Auswirkungen unterschiedlicher Navigationsassistenz auf die räumliche Orientierung älterer Erwachsener in einer virtuellen Umgebung

Mode Awareness im Fahrkontext: Eine theoretische Betrachtung

Maria Kolbig¹ und Stephan Müller¹

Schlüsselwörter: mode awareness, Fahrerassistenzsysteme, situation awareness, Informationsverarbeitungsprozess, Mensch-Maschine-Systeme.

Zusammenfassung

Die Einordnung des Konstrukts mode awareness in den Fahrkontext gewinnt zunehmend an Relevanz aufgrund wachsender Komplexität im Bereich der Fahrerassistenzsysteme. Ein umfassendes theoretisches Verständnis von mode awareness bildet die Grundlage für zielgerichtete Gestaltung von Fahrerassistenzsystemen hinsichtlich Akzeptanz und Sicherheit. Auf Basis bestehender Definitionen anderer Forschungsbereichen wird eine neue Definition im Fahrkontext erarbeitet. Desweiteren wird das Konstrukt mode awareness vom bekannten Konstrukt situation awareness (Endsley, 1995; Rauch, 2006) abgegrenzt und mit bestehenden Theorien weiterentwickelt. Somit folgt eine Einordnung in den menschlichen Informationsverarbeitungsprozess und eine Identifikation der wichtigsten Einflussfaktoren auf mode awareness. Diese Einflussfaktoren werden im Folgenden auf den Fahrkontext übertragen und durch fahrkontextspezifische Faktoren ergänzt. Auf Basis dieser Erkenntnisse können Ansätze zur Beeinflussung von mode awareness abgeleitet werden. In anschließenden Schritten kann dieses theoretische Konstrukt hinsichtlich verschiedener Einflussfaktoren empirisch untersucht werden. Die Untersuchung von mode awareness hat u.a. die Ableitung systemspezifische Gestaltungsempfehlungen zum Ziel.

Einleitung

Ende des 20. Jahrhunderts hat der Begriff mode awareness seinen Ursprung in Untersuchungen der Luftfahrtbranche zu situation awareness. Dabei stellte sich heraus, dass es Probleme mit mode awareness geben kann, wenn automatisierte Systeme dem Piloten seine primäre Aufgabe teilweise oder ganz abnehmen (Sarter, 1994). Seitdem wird der Begriff in vielen Forschungsbereichen der Mensch-System-Interaktion (u.a. Medizintechnik, Mensch-Computer-Interaktion) verwendet, wenn automatisierte Systeme die Aufgaben der Benutzer übernehmen und dabei überwacht werden müssen. Die Erkenntnisse der verschiedenen Forschungsbereiche verdeutlichen die Wichtigkeit von mode awareness für eine fehlerminimierte Mensch-System-Interaktion bei zunehmender Automatisierung (Norman, 1990; Sarter & Wood, 1995; Andre & Degani, 1997). Das Konstrukt mode awareness spielt auch im Fahrkontext aufgrund zunehmender Automatisierung der primären

¹ BMW Group, Abteilung Anzeige- und Bedienkonzepte, München

Fahraufgabe durch die Einführung und Weiterentwicklung von Fahrerassistenzsystemen (Kompaß, 2008) eine wichtige Rolle (König, Weiß & Mayser, 2002). Sicherheitskritische, akzeptanz- und vertrauensschmälernde Effekte von Fahrerassistenzsystemen können mit Untersuchungen der mode awareness frühzeitig erkannt und vermieden werden (Morgan & Hancock, 2011).

Aufgrund bestehender Unterschiede hinsichtlich des Umfelds, der psychologischen Gegebenheiten der Benutzer und der Auslegung der automatisierten Systeme in den verschiedenen Interaktionsbereichen müssen die Erkenntnisse auf ihre Übertragbarkeit auf den Automobilbereich geprüft werden. Zunächst wird mode awareness definiert, um anschließend den Aufbauprozess und dessen Einflussfaktoren besser analysieren zu können.

Das Konstrukt mode awareness

Eine häufig zitierte Definition von Sarter und Woods (1995) aus der Luftfahrt bezeichnet Benutzerwissen als mode awareness, welches einen benötigten Umfang an Informationen über das Systemverhalten enthält. Zu diesem Wissen zählen nicht nur die Systemzustände und die dazugehörigen Verhaltensweisen des Systems, sondern auch einzelnen Parameter, die diese Verhaltensweisen charakterisieren. Bezieht man die Definition für awareness von Klatzky (1984) mit ein, wird deutlich, dass mode awareness eine ausreichende Qualität des Wissens über das automatisierte System für fehlerminimierte Interaktion bedeutet. Allgemein gefasst bezeichnet mode awareness somit die korrekte Verknüpfung aller benötigten Informationen im mentalen Modell in Abhängigkeit der Aufgabe (Abb.1).

Weißt das mentale Modell nicht Lücken für die Aufgabenbewältigung auf, so spricht Dekker (2006) von automation surprise (Abb.1). Dies wird in der Literatur unterschieden in mode confusion, bei der benötigte Informationen fehlen (Bredereke & Lankenau, 2005), und mode error, bei dem vorhandene Informationen und deren Verknüpfungen fehlerhaft sind (Dekker, 2006).

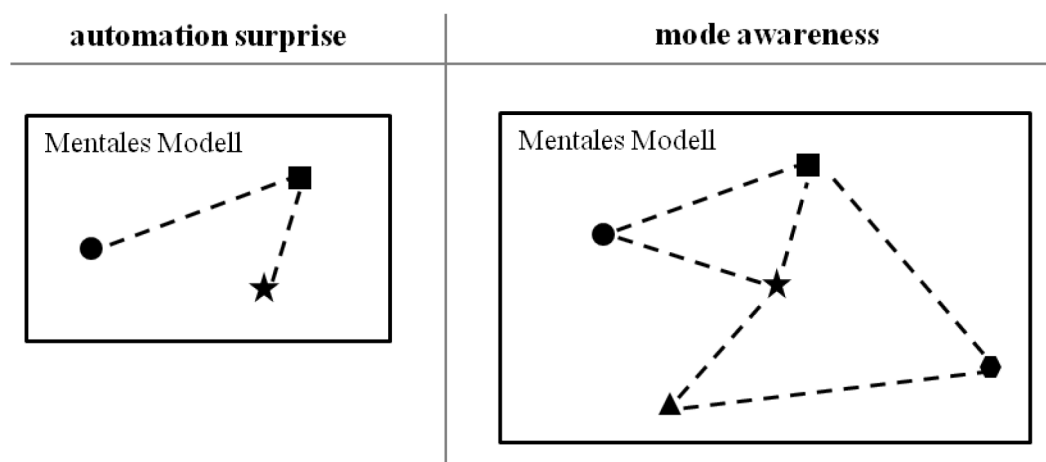


Abb.1: Gegenüberstellung des Informationsumfangs eines mentalen Modells bei automation surprise und mode awareness.

Auffällig ist die Ähnlichkeit der Definition für mode awareness mit der Definition für situation awareness von Endsley (1995), wodurch eine Verbindung nahe gelegt wird, dass mode awareness eine bestimmte Ausprägung von situation awareness ist. Diese Verbindung der beiden Konstrukte scheint sich auch bei Wickens wiederzufinden (Buld & Krüger, 2002). Somit wird hier die Annahme getroffen, dass ein Erreichen von mode awareness auf einem gleichen Prozess wie situation awareness basiert. Der Unterschied zwischen den beiden Konstrukten ist der Informationsumfang der mentalen Modelle.

Während sich der Informationsumfang bei mode awareness ausschließlich auf das System umfasst, sind für situation awareness zudem alle Informationen über die momentane Situation, in der sich System und Benutzer befinden, von Bedeutung (Abb.2).

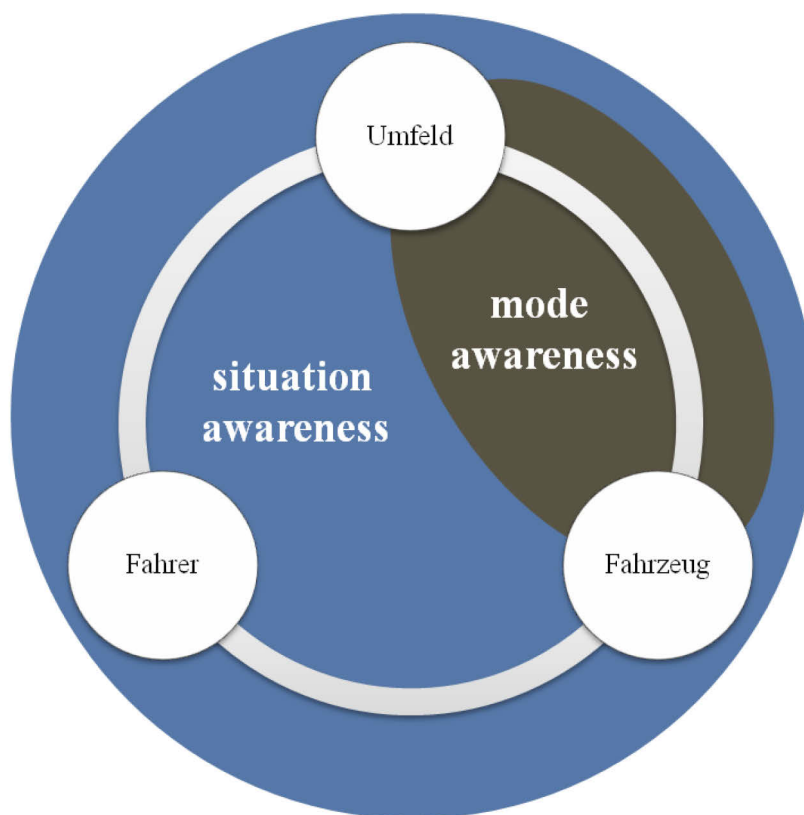


Abb.2: Mode awareness als Teil von situation awareness im Fahrkontext.

Überträgt man den bekannten Informationsverarbeitungsprozess von situation awareness mit den drei Schritten Informationsaufnahme, -interpretation und -projektion auf mode awareness, dann sind die Eingangsgrößen dieses Prozesses Systeminformationen, systemrelevante Umfeldinformationen und Systemerfahrung. Diese drei Eingangsgrößen kommen auch im Fahrkontext vor (Norman, 2011), dabei beziehen sich die Systemrückmeldungen nur auf Fahrerassistenzsysteme (Abb.3).

In einem mentalen Modell werden diese Eingangsdaten verknüpft. Somit ist die Voraussetzung für die Verständnisbildung bezüglich Systemzweck, -

funktion, -zustand und -zustandsprognose gegeben (Rouse & Morris, 1986; Rushby, 2001) – also für mode awareness. Zumeist sind diese Informationen dem Systembenutzer nicht vollständig bewusst – einige davon stehen nur unbewusst zur Verfügung. Laut Kihlstrom (1984) bilden sowohl bewusste als auch unbewusste Information awareness (Abb.3).

Aus der Luftfahrtbranche ist bekannt, dass für eine lückenlose Kenntnis relevanter Systeminformationen die Systemrückmeldungen eine wichtige Rolle spielen und somit eine Auslegung der Rückmeldungen zur Unterstützung von mode awareness von großer Bedeutung sind (Sarter & Woods, 1995). Dieser Grundgedanke kann mit Anpassungen auf den Fahrkontext übertragen werden.

Bei Piloten besteht die primäre Informationsquelle zumeist nur aus den Cockpitanzeigen – im Fahrkontext erstreckt diese sich weit über den Fahrzeuginnenraum hinaus. Piloten nehmen teilweise weniger und zeitunkritischere Informationen auf als Fahrer, woraus sich eine höhere Informationsrate im Fahrkontext ergibt (Gish & Staplin, 1995). Der zeitliche Rahmen, der dem Fahrer zur Wahrnehmung der Informationen zur Verfügung steht wird durch die momentane Fahrzeugführungsebene definiert (Thompson, 2005). Demgegenüber steht die Leistungsfähigkeit des Fahrers. Fahrer unterliegen im Gegensatz zu Piloten nur geringen Auswahlkriterien hinsichtlich ihrer visuellen und kognitiven Leistungsfähigkeit. Darum variiert das Spektrum dieser Leistungsfähigkeit bei Fahrern enorm, was bei der Auslegung der Systemrückmeldungen unbedingt beachtet werden sollte.

Für die Begrenzung dieser Leistungsfähigkeit spielen allgemein vier Faktoren eine wichtige Rolle: Zunächst ist angesichts der bereits hohen visuellen Belastung des Fahrers bei Bewältigung der Fahraufgabe eine Überforderung durch die Summe aller Systemrückmeldungen zu vermeiden. Nach Wickens (2000) sind die Sinneskanäle des Menschen hinsichtlich ihrer Leistungsfähigkeit begrenzt, so dass eine Aufteilung der Informationsübertragung an den Fahrer auf verschiedene Sinneskanäle (sogenannter Ansatz der Multimodalität) unter Umständen hilfreich sein kann. Desweiteren unterliegt das Arbeitsgedächtnis Kapazitätsbegrenzungen (Miller, 1956; Cowan, 2005), welche in Abhängigkeit der Erfahrung variieren. Miller (1956) fand heraus, dass mit steigender Aufgabenerfahrung eine effizientere Nutzung der Kapazitäten einhergeht. Das bedeutet, dass Erfahrene im Vergleich mit Unerfahrenen trotz gleicher Kapazitäten eine größere Menge an Informationen im Arbeitsgedächtnis verarbeiten können. Der dritte Aspekt ist, dass Ablenkung relevante Informationen im Arbeitsgedächtnis durch aufgabenirrelevante Informationen ersetzen oder einen Abbruch des Aufbauprozesses für ein mentales Modell herbeiführen kann (Baumann, Petzoldt, & Krems, 2006). Schließlich kommt die Theorie von Yerkes und Dodson (1908) hinzu, die besagt, dass die Leistungsfähigkeit von der physiologischen Aktivierung abhängig ist. Bei Unterforderung aber auch Überforderung nimmt die Leistungsfähigkeit ab.

Diese Faktoren kommen so auch im Fahrkontext vor und werden hier aufgrund ihres Einflusses auf mentale Modelle auch als wichtige Einflussfaktoren auf die mode awareness identifiziert.

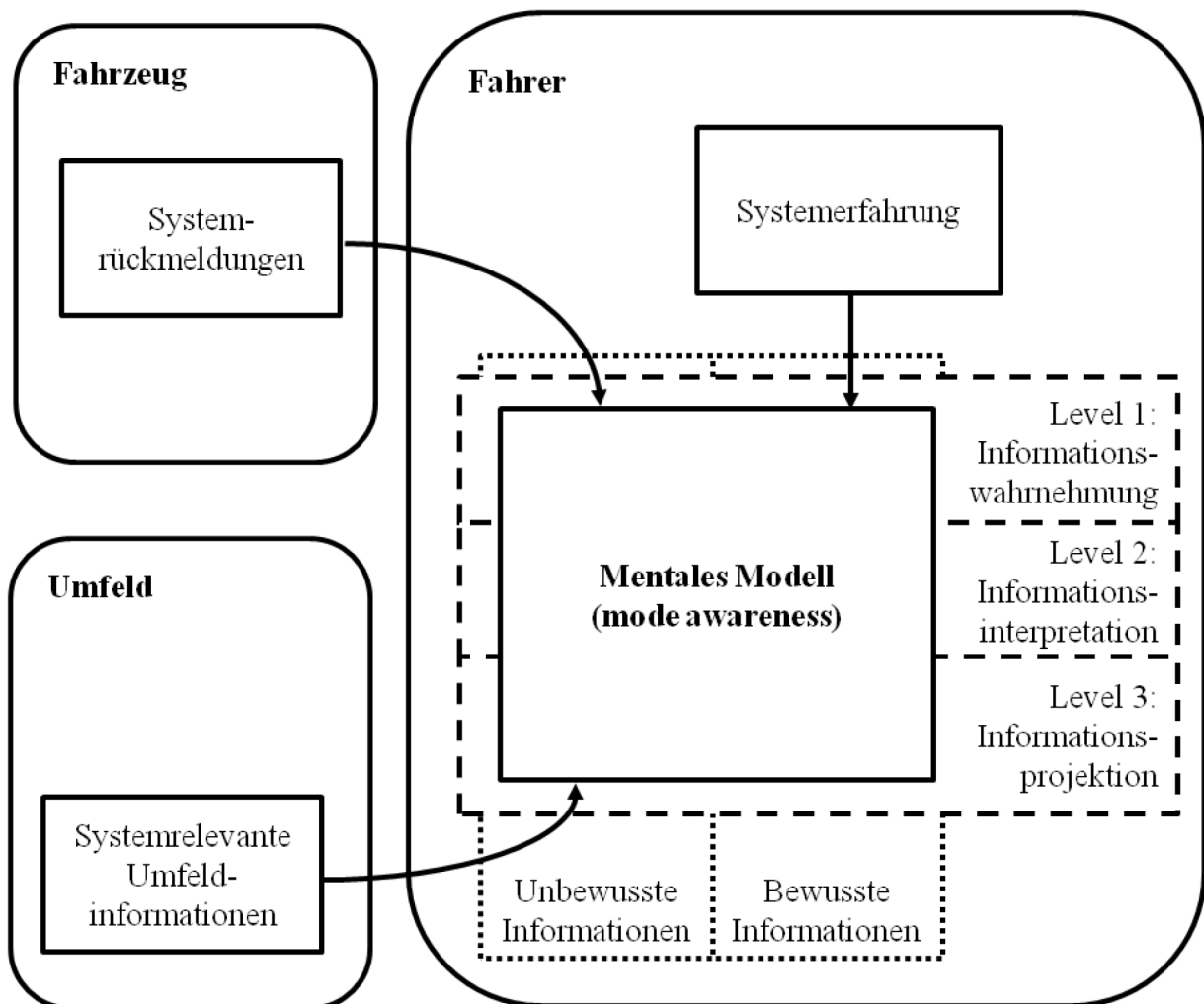


Abb.3: Entstehungsprozess von mentalen Modellen für das Erreichen von mode awareness im Fahrkontext (in Anlehnung an Endsley (1995), Kihlstrom (1984) und Norman (2011)).

Die für mode awareness optimierte Systemgestaltung wird insofern wichtig, wenn in einem Automobil verschiedene Automatisierungsgrade von Fahrerassistenzsystemen zum Einsatz kommen. Es liegen verschiedene Ansätze für die Klassifikation von Automatisierungsgraden vor (Skottke, 2007). Die Gemeinsamkeit dieser Ansätze liegt in der Annahme, dass steigende Automatisierung mit erhöhtem Funktionsumfang des Systems einhergeht, welcher sich aus der zunehmenden Übernahme der primären Fahraufgabe ergibt (z.B. Tempomat → ACC → Kreuzungsassistent → Stauassistent). Damit der Fahrer die so entstandene Überwachungsaufgabe hinreichend erfüllen kann, ist mode awareness notwendig. Voraussetzung dafür ist ein ausreichend umfangreiches mentales Modell über das System. Das mentale Modell wird maßgeblich durch die Systemrückmeldung gebildet. Dabei kann angenommen werden, dass zur Erreichung von mode awareness mit steigender Automatisierung auch ein steigender Rückmeldungsumfang erforderlich ist.

Da zudem die Dynamik und Komplexität der gesamten Situation sowohl das mentale Modell, wie auch den Überwachungsaufwand beeinflusst, ist darauf zu achten, dass der Fahrer zu keinem Zeitpunkt überfordert wird.

Die oben genannten allgemeinen und fahrkontextspezifischen Gegebenheiten müssen bei der Gestaltung von Systemrückmeldungen beachtet werden, um den Aufbau eines mentalen Modells hinsichtlich der Erreichung von mode awareness zu unterstützen.

Ausblick

Mode awareness definiert einen spezifischen Informationsumfang. Dieser Umfang setzt sich aus den benötigten Informationen über Systemzustände und dazugehörigen Systemverhaltensweisen zusammen und umfasst bewusste und unbewusste Informationen als Basis eines mentalen Modells. Mode awareness wird in dieser Theorie auch als ausreichend gute Qualität des mentalen Modells im Umgang mit automatisierten Systemen wie Fahrerassistenzsystemen in einem dynamischen Umfeld definiert.

Die Auslegung der Systemrückmeldungen scheint nach Betrachtung der Eingangsgrößen und wichtigsten Einflussfaktoren ein vielversprechender Ansatz zur Unterstützung von mode awareness im Fahrkontext. Bei der Auslegung spielen sowohl der Informationsgehalt und deren Kodierung als auch der Rückmeldungsart und -zeitpunkt eine entscheidende Rolle. In den nächsten Schritten muss nun die Anwendbarkeit bestehender Erkenntnisse bezüglich der vier Teilaspekte der Rückmeldungsgestaltung im Fahrkontext im Zusammenhang mit dem Konstrukt mode awareness geprüft werden. Darauf basierend können weiterführende Studien den Einfluss dieser vier Teilaspekte auf mode awareness untersuchen, um weitere geeignete Auslegungen der Rückmeldungen abzuleiten.

Literatur

- Andre, A. & Degani, A. (1997): Do You Know What Mode You're in? An Analysis of Mode Error in Everyday Things. In: Mouloua, M. & Koonce, J. M. (Hrsg.): *Human-automation interaction: Research and practice* (S. 19–28). Mahwah, NY: Lawrence Erlbaum.
- Baumann, M. & Petzoldt, T. & Krems, J. F. (2006). Situation Awareness beim Autofahren als Verstehensprozess. *MMI-Interaktiv*, 11, 43–57.
- Bredereke, J. & Lankenau, A. (2005). Safety-Relevant Mode Confusions. Modeling and Reducing Them. *Reliability Engineering & System Safety*, 88(3), 229–245.
- Buld, S. & Krüger, H.-P. (2002). Wirkung von Assistenz und Automation auf Fahrerzustand und Fahrsicherheit. Projekt EMPHASIS (Effort-Management and Performance Handling in sicherheitsrelevanten Situationen) (19 S 9812 7). Würzburg: Interdisziplinäres Zentrum für Verkehrswissenschaften.

- Cowan, N. (2005). Working memory capacity. Essays in Cognitive Psychology. New York, NY: Psychology Press.
- Dekker, S. (2006). The field guide to understanding human error. Aldershot: Ashgate.
- Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *The Journal of the Human Factors and Ergonomics Society*, 37(1), 32–64.
- Gish, K. W. & Staplin, L. (1995). Human Factors Aspects of Using Head Up Display in Automobiles: A Review of Literature (Interim Report, DOT HS 808 320). Washington, D.C.: U. S. Department of Transportation.
- Kihlstrom, J. F. (1984). Conscious, Subconscious, Unconscious: A Cognitive Perspective. In: Bowers, K. S. & Meichenbaum, D. (Hrsg.): *The unconscious reconsidered* (pp. 149–211). New York, NY: Wiley–Interscience.
- Klatzky, R. L. (1984). *Memory and awareness. An information-processing perspective*. New York, NY: W.H. Freeman and Company.
- Kompaß, K. (2008). Fahrerassistenzsysteme der Zukunft - auf dem Weg zum autonomen Pkw? In: Schindler, V. & Sievers, I. (Hrsg.): *Forschung für das Auto von morgen. Aus Tradition entsteht Zukunft* (S. 261–286). Berlin: Springer.
- König, W. & Weiß, K.-E. & Mayser, C. (2002). S.A.N.T.O.S. Situations – Angepasste und Nutzer – Typ-zentrierte Optimierung von Systemen zur Fahrerunterstützung (Projektabschlussbericht). München: Robert Bosch GmbH, BMW Group.
- Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two. Some Limits on Our Capacity for Processing Information. *Psychological Review*, 63, 81–97.
- Morgan, J. F. & Hancock, P. A. (2011). The Effect of Prior Task Loading on Mental Workload: An Example of Hysteresis in Driving. *The Journal of the Human Factors and Ergonomics Society*, 53(1), 75–86.
- Norman, D. A. (1990). The "problem" of automation: Inappropriate feedback and interaction, not "over-automation". *Philosophical Transactions of the Royal Society of London*, B 327, 585–593.
- Norman, D. A. (2011). *Living with complexity*. Cambridge: MIT Press.
- Rauch, N. (2009). Ein verhaltensbasiertes Messmodell zur Erfassung von Situationsbewusstsein im Fahrkontext. Dissertation, Julius-Maximilians-Universität Würzburg. Würzburg.
- Rouse, W. B. & Morris, N. M. (1986). On Looking into the Black Box. Prospects and Limits in the Search for Mental Models (Technical Report, 85–2). Atlanta: Office of Naval Research.
- Rushby, J. (2002). Using model checking to help discover mode confusions and other automation surprises. *Reliability Engineering and System Safety*, 75(2), 167–177.

- Sarter, N. & Woods, D. (1994). Pilot interaction with cockpit automation: II. An experimental study of pilots' model and awareness of the Flight Management System. *International Journal of Aviation Psychology*, 4, 1–28.
- Sarter, N. & Woods, D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *The Journal of the Human Factors and Ergonomics Society*, 37(1), 5–19.
- Skottke, E.-M. (2007). Automatisierter Kolonnenverkehr und adaptiertes Fahrverhalten. Untersuchung des Abstandsverhaltens zur Bewertung möglicher künftiger Verkehrsszenarien. Dissertation, Rheinisch-Westfälische Technische Hochschule Aachen. Aachen:
- Thompson, L. (2005). Entwicklung und Evaluierung eines Anzeige-Bedienkonzepts für Fahrerassistenz auf der Fahrzeugführungsebene. Masterarbeit, Technische Universität München, München.
- Wickens, C. D. & Hollands, J. G. (2000). *Engineering psychology and human performance* (3. Aufl.). Upper Saddle River, NJ: Prentice-Hall.
- Yerkes, R. M. & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459–482.

Evaluation eines haptischen Touchpads für die Fahrer-Fahrzeug-Interaktion

Michael Domhardt¹, Ercan Tunca², Ingo Zoller², Peter Lotz² und Ludger Schmidt¹

Schlüsselwörter: Touchpad, haptische Rückmeldung, Lane Change Test, Fahrer-Fahrzeug-Interaktion, Emotionsanalyse

Zusammenfassung

In aktuellen Fahrzeugen sind teilweise bereits Touchpads als Nachfolger von Dreh-Drück-Stellern, mechanischen Tasten und Schaltern zu finden. Doch im Gegensatz zu ihren mechanischen Vorgängern fehlt bei der Touchpadinteraktion eine haptische Rückmeldung bei der Auswahl und Betätigung eines auf einem Bildschirm dargestellten Eingabeelements, wie zum Beispiel einer Taste. Um diesen Nachteil zu beheben, wurde von der Firma Continental der Prototyp eines Touchpads entwickelt, der auf Basis einer eindimensional angeregten Masse haptische Rückmeldungen ermöglicht. Im Rahmen einer empirischen Nutzerstudie wurde der Einfluss der elektromechanischen haptischen Rückmeldung bei der Fahrer-Fahrzeug-Interaktion auf Fahrerablenkung, Effektivität, Effizienz des Touchpads und Zufriedenstellung des Fahrers untersucht. Die Evaluation des Prototyps des haptischen Touchpads wurde mit 32 Probanden im Fahrsimulator durchgeführt. Mit Hilfe des Lane Change Tests, Videobeobachtung und automatisierter Emotionsanalyse konnte ein signifikant positiver Einfluss für die Fahrerablenkung und die Effizienz bei der Bewältigung von Nebenaufgaben festgestellt werden. Die Ergebnisse zur Effektivität und Zufriedenstellung zeigten keinen signifikanten Einfluss.

Problemstellung und Zielsetzung

Die Fahrer-Fahrzeug-Interaktion hat sich in den letzten Jahren im Bereich der Nebenaufgaben von der Eingabe mittels physischer Schalter und Taster zur Interaktion mit grafischen Benutzungsoberflächen weiterentwickelt. Zur Eingabe dienen neben Dreh-Drück-Stellern und Touchscreens auch Touchpads. Die beiden letztgenannten Eingabemittel verhindern bei fehlender haptischer Rückmeldung die Blindbedienung, die im Fahrzeug wünschenswert wäre, um die Ablenkungswirkung zu minimieren. Bei der Verwendung von Touchscreens besteht zusätzlich das Problem, dass sich der Bildschirm durch die Anbringung im Reichraum lediglich im Blickfeld statt im Gesichtsfeld befindet und die Augen durch die notwendige Akkommodation zwischen entfernter Straße und nahem Bildschirm zusätzlich beansprucht werden.

¹ Universität Kassel, Fachgebiet Mensch-Maschine-Systemtechnik

² Continental Automotive GmbH, Business Unit Instrumentation & Driver HMI

Zur künstlichen Erzeugung einer haptischen Rückmeldung werden bisher zwei Ansätze verfolgt. Zum einen werden haptische Rezeptoren mittels elektromechanischer Aktoren erregt, was einem adäquaten Reiz entspricht. Technisch ist dies über exzentrisch rotierende Massen, lineare Resonanzaktuatoren, Piezomodule oder elektroaktive Polymeraktoren realisierbar (Immersion 2012). Aktuell sind mehrere Systeme in Entwicklung (z.B. Blattner et al. 2012, Blattner et al. 2012 a, Pitts et al. 2012). Auch über das Prinzip des Ladungsträgeraustauschs können die Haptikrezeptoren erregt werden (DIN EN ISO 9241-910). Entsprechende Systeme sind bisher nur aus wenigen Laborstudien bekannt (z. B. Bau et al. 2010) und kommerziell noch nicht verfügbar.

Für die ergonomische Interaktion mit dem Infotainmentsystem im Fahrzeug wurde von der Firma Continental ein elektromechanischer Ansatz verfolgt und der Prototyp eines Touchpads entwickelt, der auf Basis einer eindimensional angeregten Masse haptische Rückmeldungen ermöglicht (Abb. 1).



Abb. 1: Touchpad mit elektromechanischer haptischer Rückmeldung

Das entwickelte Interaktionskonzept mit haptischer Rückmeldung soll im Fahrzeug zur Bewältigung von fahrzeugtypischen Nebenaufgaben eingesetzt werden. Beispiele für solche Nebenaufgaben sind die Einstellung der Klimaanlage oder die Eingabe eines Ziels ins Navigationssystem. Bei der Bewältigung von Nebenaufgaben werden perzeptive, kognitive und motorische Ressourcen des Fahrers benötigt, die dann für die Fahrzeugführung nicht zur Verfügung stehen. Eine Nebenaufgabe stellt daher immer eine Ablenkung von der eigentlichen Fahraufgabe dar. Ein Touchpad mit haptischer Rückmeldung bietet das Potenzial, diese Ablenkung zu verringern. Das Ziel dieser Studie war daher die wissenschaftlich fundierte Beurteilung der Wirksamkeit dieses neuartigen Interaktionskonzepts.

Versuchsdesign und Versuchsdurchführung

Um den Einfluss der haptischen Rückmeldung auf die Fahrer-Fahrzeug-Interaktion wissenschaftlich beurteilen zu können, wurde ein Laborversuch im Fahrsimulator konzipiert. Die Aufgabe der Fahrzeugführung wurde durch eine Spurwechselaufgabe nach ISO 26022 realisiert. Abb. 2 zeigt ein Beispiel der Außenansicht, bei der der Proband gemäß den Schildern in die rechte Spur wechseln sollte.

Parallel zur Fahraufgabe sollten Nebenaufgaben mit dem haptischen Touchpad absolviert werden. Die zum Touchpad gehörende grafische Benutzungsschnittstelle stellte im Ausgangszustand auf einem kleinen Bildschirm in der Mittelkonsole ein

Hauptmenü dar (Abb. 3 links) und ermöglichte im Untermenü Control vier unterschiedliche fahrzeugtypische Nebenaufgaben. Es war jeweils das Betätigen der Heckscheibenheizung, der Frontscheibenheizung, der Umluft und der Klimaanlage möglich (Abb. 3 rechts).



Abb. 2: Fahrsimulation der Spurwechselaufgabe nach ISO 26022

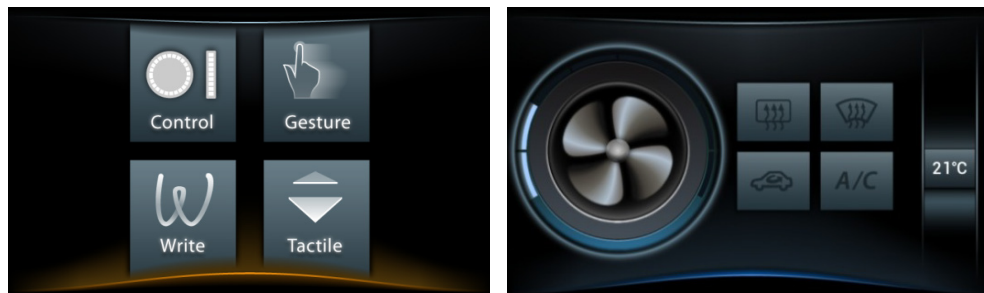


Abb. 3: Hauptmenü (links) und Untermenü Control (rechts) der Benutzungsschnittstelle

Als unabhängige Variable wurde der Zustand der elektromechanischen haptischen Rückmeldung bei der Interaktion mit dem Touchpad (ein- bzw. ausgeschaltet) festgelegt. Die abhängigen Variablen der Untersuchung waren die Fahrerablenkung sowie die Gebrauchstauglichkeit in den Ausprägungen Effektivität, Effizienz und Zufriedenstellung. Die Fahrerablenkung wurde gemäß ISO 26022 anhand der Fahrleistung beim Ausführen der genormten Spurwechselaufgabe sowie der Blickzuwendung zur Nebenaufgabe ermittelt. Die Effizienz und Effektivität des Touchpads ergaben sich aus der erfassten Nutzerinteraktion beim Lösen der Nebenaufgaben. Zur Ermittlung der Zufriedenstellung wurde eine softwaregestützte Emotionsanalyse der Fahrermimik nach Ekman (1970) bei der Fahrer-Fahrzeug-Interaktion vorgenommen, um den Versuch nicht durch eine Befragung unterbrechen zu müssen. Die abhängigen Variablen mit der jeweiligen Operationalisierung sind in Tab. 1 zusammengefasst.

Tab. 1: Operationalisierung der abhängigen Variablen

Abhängige Variable	Operationalisierung
Fahrerablenkung	Fahrleistung und Blickzuwendung
Effektivität	Aufgabenerfüllung
Effizienz	Fehlerrate und Aufgabendauer
Zufriedenstellung	positive Emotionen des Fahrers

Die Arbeitshypothesen zum Untersuchungsgegenstand des haptischen Touchpads gemäß der definierten Variablen lauteten wie folgt:

- H₁ Die Fahrerablenkung ist mit elektromechanischer haptischer Rückmeldung geringer als ohne elektromechanische haptische Rückmeldung.
- H₂ Die Effektivität ist mit elektromechanischer haptischer Rückmeldung höher als ohne elektromechanische haptische Rückmeldung.
- H₃ Die Effizienz ist mit elektromechanischer haptischer Rückmeldung besser als ohne elektromechanische haptische Rückmeldung.
- H₄ Die Zufriedenstellung ist mit elektromechanischer haptischer Rückmeldung besser als ohne elektromechanische haptische Rückmeldung.

Daraus ergaben sich die entsprechenden Nullhypothesen zu:

- H₀₁ Die Fahrerablenkung ist unabhängig von der elektromechanischen haptischen Rückmeldung.
- H₀₂ Die Effektivität ist unabhängig von der elektromechanischen haptischen Rückmeldung.
- H₀₃ Die Effizienz ist unabhängig von der elektromechanischen haptischen Rückmeldung.
- H₀₄ Die Zufriedenstellung ist unabhängig von der elektromechanischen haptischen Rückmeldung.

Die Versuchsumgebung für den Fahrversuch umfasste neben einem Fahrzeug-Mockup mit Fahrersitz, Pedalerie und Lenkrad das haptische Touchpad in der Mittelarmlehne und einen Bildschirm in der Mittelkonsole (Abb. 4). Zur Beobachtung des Versuchs kamen drei Kameras zum Einsatz. Eine Kamera erfasste die seitliche Totale und je eine weitere Kamera diente zur Erfassung des Gesichts aus Fahrt- und Bildschirmrichtung. Zum Betrieb des Versuchsaufbaus waren drei vernetzte Computer notwendig, die über einen netzwerkinternen Zeitserver synchronisiert wurden.

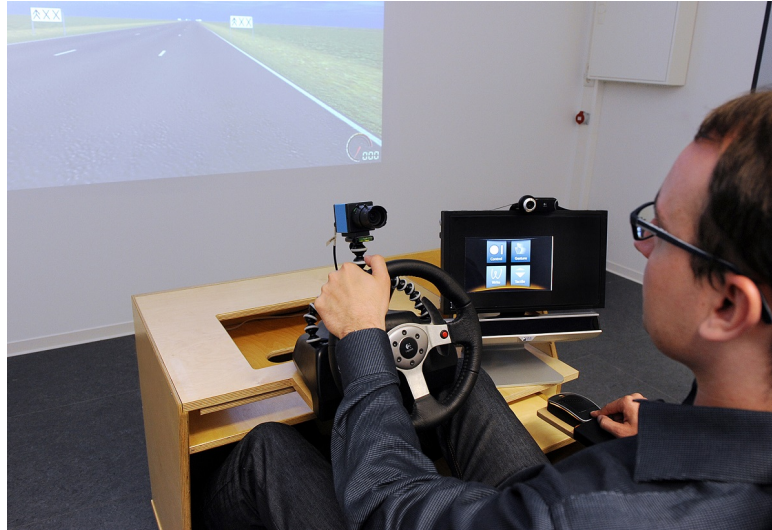


Abb. 4: Versuchsumgebung im Labor

Vor der Versuchsdurchführung wurden von den potentiellen Probanden per Onlinebefragung Alter, Geschlecht, Kontaktmöglichkeiten, der Besitz einer PKW-Fahrerlaubnis, die jährliche Fahrpraxis sowie die Händigkeit nach dem Edinburgh Inventory (Oldfield 1971) und die Technikaffinität mit elektronischen Geräten (TA-EG) nach Karrer et al. (2009) erfasst. Alle Umfrageteilnehmer mit einer Fahrerlaubnis wurden zur Versuchsteilnahme eingeladen. Unter ihnen wurden drei Buchgutscheine verlost.

Die Durchführung des Versuchs dauerte im Schnitt je Proband eine halbe Stunde. Zur Begrüßung des Teilnehmers wurde ihm eine schriftliche Information zu Ethik und Datenschutz ausgehändigt. Fahrersitz, Mittelarmlehne und Kameras wurden auf die Körpermaße des Probanden eingestellt. Nach dem Start der Videoaufzeichnung fand eine Übungsrunde mit dem Fahrsimulator statt. Dabei wurden die Probanden bereits mit der Spurwechselaufgabe konfrontiert, bei der sie eine quasirandomisierte Spurwechselserie möglichst präzise absolvieren mussten. Es folgte die Einführung des Probanden in die Interaktion mit dem haptischen Touchpad und der grafischen Benutzungsschnittstelle. Dazu wurde der Proband aufgefordert, die Funktionen des Touchpads auszuprobieren. Die Menüs und Schaltflächen wurden vom Versuchsleiter erklärt.

Nachdem letzte Fragen geklärt waren, wurde die Referenzfahrt der Spurwechselaufgabe durchgeführt. Direkt im Anschluss folgten die beiden Fahrten, bei denen mit dem Touchpad Nebenaufgaben zu bewältigen waren. Die Nebenaufgaben umfassten jeweils drei Interaktionsschritte, die mit dem Touchpad durchgeführt wurden: Zuerst musste ausgehend vom Hauptmenü aus vier Möglichkeiten das Untermenü Control gewählt werden (Abb. 3 links). In diesem Untermenü war eine von vier Schaltflächen zu betätigen (Abb. 3 rechts). Der letzte Interaktionsschritt umfasste das Betätigen einer dedizierten Taste, die zurück ins Hauptmenü führte. In Abhängigkeit einer geraden oder ungeraden Probandennummer wurde entweder der erste oder der zweite Fahrdurchgang mit Nebenaufgaben mit elektromechani-

scher haptischer Rückmeldung durchgeführt. Die jeweils andere Fahrt fand ohne haptische Rückmeldung statt.

Auswertung und Ergebnisse

Die Stichprobe bestand aus 20 männlichen und 12 weiblichen Probanden. Das Durchschnittsalter lag bei 41,6 Jahren (Standardabweichung: 18,8 Jahre). Die Fahrpraxis der Stichprobe beträgt im Mittel 13 Tkm pro Jahr (Standardabweichung: 11 Tkm pro Jahr). Die erfragte Technikaffinität der Stichprobe nach Karrer et al. (2009) ist in Abb. 5 abgebildet. Der Standardfragebogen umfasst vier Subskalen: Begeisterung für Technik, Kompetenz im Umgang mit Technik, positive Technikfolgen und negative Technikfolgen. Items mit einem negativen Einfluss auf die Technikaffinität werden bei der Auswertung gepolt, so dass auf jeder Skala Werte zwischen eins (geringe Technikaffinität) und fünf (hohe Technikaffinität) erreichbar sind. Die Auswertung des Edinburgh Inventory nach Oldfield (1971) ergab einen Links- und 31 Rechtshänder.

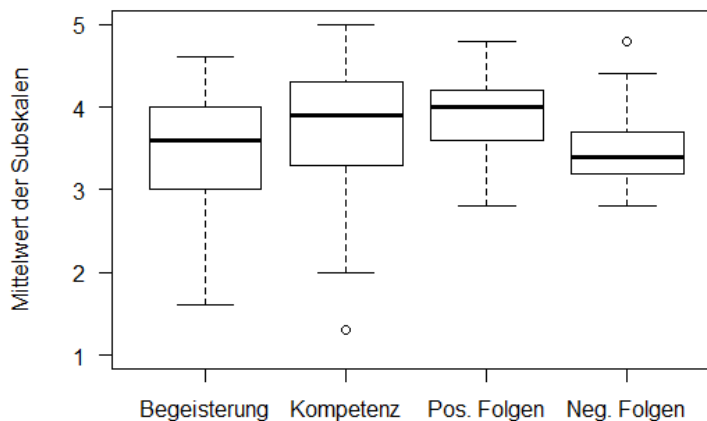


Abb. 5: Technikaffinität der Stichprobe (n = 32)

Die Fahrleistung im Lane Change Test ergibt sich nach ISO 26022 aus der gemittelten Spurbabweichung beim Fahren mit gleichzeitiger Bewältigung von Nebenaufgaben von der individuellen Solltrajektorie aus der Referenzfahrt (Abb. 6 links). Da aufgrund der Stichprobeneigenschaften keine Normalverteilung angenommen werden kann, wurde ein Wilcoxon-Rangsummentest durchgeführt. Die Ausprägung der unabhängigen Variablen hat keinen signifikanten Einfluss auf die Fahrleistung beim Lane Change Test ($p = 0,33$).

Zur Ermittlung der Blickzuwendung wurden die beiden Videos vom Gesicht kodiert. Die Blickzuwendung zum Bildschirm in der Mittelkonsole während der Bearbeitung der Nebenaufgaben wurde für jeden Probanden pro Nebenaufgabe ermittelt (Abb. 6 rechts). Mit elektromechanischer haptischer Rückmeldung fällt die Blickdauer im Mittel um 23 % kürzer aus ($p < 0,01$). Die elektromechanische haptische Rückmeldung hat folglich einen signifikant positiven Einfluss auf die Fahrerablenkung (H_{01} verworfen, H_1 angenommen).

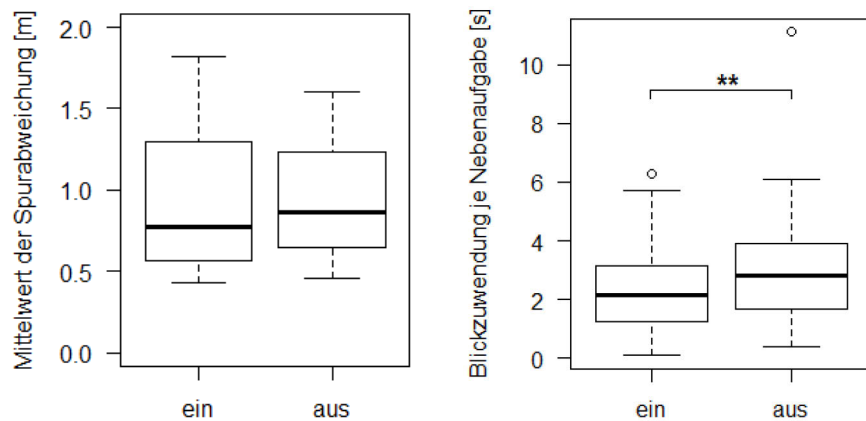


Abb. 6: Fahrleistung (links) und Blickzuwendung zum Mittelkonsolenbildschirm (rechts) nach Zustand der elektromechanischen haptischen Rückmeldung (ein/aus, n = 32)

Die Effektivität ergibt sich aus dem Erfüllungsgrad der Nebenaufgaben. Jeder Proband konnte aufgrund des Versuchsdesigns unabhängig von der Ausprägung der unabhängigen Variablen alle vier Nebenaufgaben erfolgreich absolvieren. Da kein Einfluss der unabhängigen Variablen feststellbar ist, wird H_{02} angenommen und H_2 verworfen.

Die Dauer zur Erfüllung einer Nebenaufgabe und die dabei verursachten Fehler lassen Rückschlüsse auf die Effizienz zu. Die unabhängige Variable hat keinen signifikanten Einfluss auf die Fehlerrate, wobei eine Tendenz für einen positiven Einfluss sichtbar ist (Abb. 7 links). Die durchschnittliche Dauer zur Bewältigung einer Nebenaufgabe sinkt durch die elektromechanische haptische Rückmeldung im Mittel um 33 % ($p = 0,02$). Daher kann H_{03} verworfen und H_3 angenommen werden.

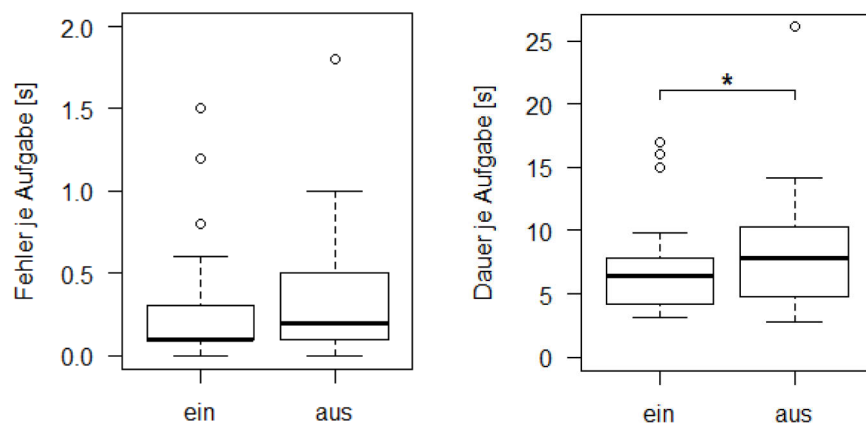


Abb. 7: Fehlerrate (links) und Aufgabendauer (rechts) nach Zustand der elektromechanischen haptischen Rückmeldung (ein/aus, n = 32)

Die automatisierte Emotionsanalyse gelang aufgrund hardwaretechnischer Einschränkungen lediglich bei elf Probanden. Auf Basis der sieben klassifizierten Grundemotionen nach Ekman (1970) wurde die Dauer der positiven Emotion (happy) für jeden Proband aufsummiert (Abb. 8). Es war kein signifikanter positiver Einfluss der unabhängigen Variable auf die Dauer der positiven Emotion

nachweisbar ($p = 0,90$). Allerdings kann eine Tendenz für den gegenteiligen Effekt beobachtet werden. Die Nullhypothese H_{04} muss daher angenommen und die Alternativhypothese H_4 verworfen werden.

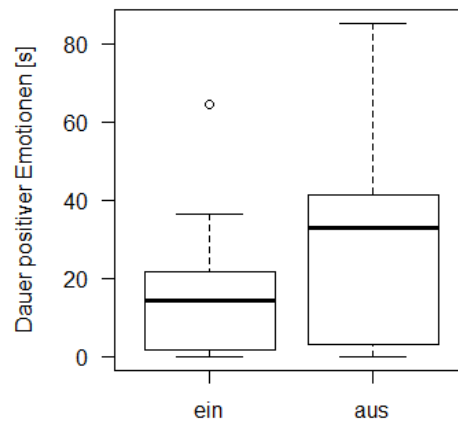


Abb. 8: Fahreremotionen nach Zustand der elektromechanischen haptischen Rückmeldung (ein/aus, $n = 11$)

Über die Prüfung der Hypothesen hinaus wurden die Daten explorativ ausgewertet. Um die Altersabhängigkeit der Ergebnisse abschätzen zu können, erfolgte eine Aufteilung der Stichprobe: die jüngsten 40 % der Stichprobe bildeten die Gruppe der jungen Probanden ($n = 13$, $\bar{x} = 25,2$ Jahre), die ältesten 40 % der Stichprobe ergaben die Gruppe der älteren Probanden ($n = 13$, $\bar{x} = 62,7$ Jahre), dazwischen lagen die mittelalten Probanden ($n = 6$). Bei der Gegenüberstellung der jungen und der älteren Probanden konnten für die jungen Probanden eine 44 % bessere Fahrleistung ($p < 0,01$), eine 33 % geringere Blickzuwendung während der Nebenaufgabenbearbeitung ($p < 0,01$), eine 38 % geringere Aufgabendauer ($p < 0,01$) und eine 60 % geringere Fehlerrate ($p < 0,01$) festgestellt werden. Das Alter der Probanden hatte demzufolge einen starken Einfluss auf die Fahrerablenkung und Effizienz.

Diskussion und Ausblick

Da in allen Subskalen der Technikaffinität drei Viertel der Probanden oberhalb der Skalenmitte liegen, kann von einer überdurchschnittlich technikaffinen Stichprobe ausgegangen werden. Die nach ISO 26022 ermittelten Werte für die Fahrleistung fallen fast identisch zu Pitts et al. (2012) aus, dessen Untersuchungsgegenstand ein Touchscreen mit elektromechanischer haptischer Rückmeldung war. Die erreichten Ergebnisse der aufsummierten Blickzuwendung zeigen, dass mit aktivierter elektromechanischer haptischer Rückmeldung die Ablenkungsdauer fast um ein Viertel reduziert werden kann. Die benötigte Dauer zur Bewältigung der gestellten Nebenaufgaben ist schlecht mit anderen Studien vergleichbar, da dieser Wert neben der individuellen Interaktionsleistung des Probanden zusätzlich von Art und Umfang der gestellten Nebenaufgaben sowie der Gestaltung der grafischen Benutzeroberfläche abhängig ist. Den bisherigen Ergebnissen steht ein gegenteiliger Effekt bei der positiven Fahreremotion gegenüber. Eine mögliche Ursache kann in

der signifikant kürzeren Dauer zur Bearbeitung der Nebenaufgaben liegen. Bei einer kürzeren Interaktion entfällt auch weniger Aufmerksamkeit auf das Touchpad und die Rückmeldung. Hier kann die gezielte Auswertung der Emotionen in Bezug auf die tatsächlichen Interaktionszeiträume zur Aufklärung verhelfen. Zusätzlich kann der Effekt auch durch das Modell der Emotionsklassifikation verursacht worden sein, da die Emotion „überrascht“ den neutralen Emotionen zugeordnet ist. Dies kann bewirken, dass sich die Zeitanteile der positiven Emotion durch die bisher ungewohnte Rückmeldung zu Gunsten der neutralen Emotionen verschieben. Der nachgewiesene Einfluss des Alters auf Fahrleistung, Blickzuwendung, Fehlerrate und Aufgabendauer macht deutlich, dass für ältere Nutzer ein höherer Unterstützungsbedarf besteht.

Um die offenen Punkte zur Fahrleistung und zur Fehlerrate klären zu können, ist eine Vergrößerung der Stichprobe mit dem vorhandenen Versuchsdesign angeraten. In weiterführenden Studien können der Einfluss von Gewöhnungseffekten auf die Fahreremotionen und die Möglichkeiten zur haptischen Rückmeldung einer Formkodierung interaktiver Elemente auf berührungsempfindlichen Oberflächen untersucht werden.

Literatur

- Bau, O., Poupyrev, I., Israr, A., & Harrison, C. (2010). TeslaTouch: electrovibration for touch surfaces. In: *UIST '10: Proceedings of the 23rd annual ACM symposium on User interface software and technology*. New York, 283–292. New York: ACM.
- Blattner, A., Spiess, R., Bengler, K., & Hamberger, W. (2012). Sichere Gestaltung des Fahrer-Arbeitsplatzes durch effiziente und intuitive Bedienung eines Fahrerinformationssystems via Touchpad mit haptischer Rückmeldung. In: Gesellschaft für Arbeitswissenschaft e. V.. *Gestaltung nachhaltiger Arbeitssysteme: 58. Kongress der Gesellschaft für Arbeitswissenschaft*. Kassel, 273–277. Dortmund: GfA-Press.
- Blattner, A., Bengler, K., & Hamberger, W. (2012). Interaktionsarten eines Touchpads mit haptischer Rückmeldung zur Bedienung eines Fahrerinformationssystems. In: VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik. 6. *Fachtagung USEWARE 2012: Mensch-Maschine-Interaktion*. Kaiserslautern, 219–228. Düsseldorf: VDI.
- DIN EN ISO 9241-910 (2011). Ergonomie der Mensch-System-Interaktion – Teil 910: Rahmen für die taktile und haptische Interaktion.
- Ekman, P. (1970). Universal facial expressions of emotion. *California Mental Health Research Digest*, 8, 151–158.
- Immersion (2012, 02.04.). *Haptics in Touch Screen Hand-Held Devices*. <http://www.immersion.com/docs/Haptics-in-Touchscreen-Hand-Held-Devices.pdf>.

- ISO 26022 (2010). Road vehicles – Ergonomic aspects of transport information and control systems – Simulated lane change test to assess in-vehicle secondary task demand.
- Karrer, K., Glaser, C., Clemens, C., & Bruder, C. (2009). Technikaffinität erfassen – der Fragebogen TA-EG. In: Lichtenstein, A., Stöbel, C., & Clemens, C.. *Der Mensch im Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme*. Berlin, 196-201. Düsseldorf: VDI.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97-113.
- Pitts, M. J., Skrypchuk, L., Wellings, T., Attridge, A., & Williams, M. A. (2012). Evaluating User Response to In-Car Haptic Feedback Touchscreens Using the Lane Change Test Advances in Human-Computer Interaction. *Advances in Human-Computer Interaction*, 1-13. doi: 10.1155/2012/598739

Nutzerakzeptanz von Car-2-X-Technologien: Welche Anforderungen machen Fahrerassistenzsysteme beliebter als andere?

Madeline Volk¹ und Dr. Ingo Totzke²

Schlüsselwörter: Fahrerassistenz, Car-2-X-Technologien, Nutzerakzeptanz

Zusammenfassung

In diesem Beitrag wird die Akzeptanz von Car-2-X-Technologien (d.h. Kommunikation zwischen Fahrzeugen bzw. zwischen Fahrzeugen und der Verkehrsinfrastruktur) durch potenzielle Nutzer dargestellt. Hierzu wurden verschiedenartige Messansätze im Rahmen des Feldversuchs des Forschungsprojekts *sim^{TD}* eingesetzt (Gruppeninterviews in Workshops vs. regelmäßige Einzelbefragungen mittels Fragebögen). Ziel dieser Studie war eine Gegenüberstellung beliebter Car-2-X-Technologien (= hohe Kaufbereitschaft, sog. TOP-Funktionen) im Vergleich zu unbeliebten Technologien (= niedrige Kaufbereitschaft, sog. FLOP-Funktionen). N = 6043 Einzelbefragungs-Fragebögen und N = 399 Probanden der Gruppeninterview-Workshops aus zwei Versuchsblöcken à acht Wochen dienen als Datengrundlage. Es wird gezeigt, dass TOP-Funktionen als hilfreicher, zuverlässiger und vertrauenswürdiger bewertet werden als FLOP-Funktionen. Die TOP-Funktionen zeichnen sich insbesondere durch einen wahrnehmbaren Mehrwert im Vergleich zu bestehenden Technologien bzw. Services aus. Die Ergebnisse werden im Zusammenhang mit den gegebenen Rahmenbedingungen des Feldversuchs diskutiert.

Abstract

In the current paper, user acceptance of Car-2-X-technologies (i.e. Car-to-Car or Car-to-Infrastructure Communication) is analyzed. For this means, a multi-method approach (group-interviews in workshops vs. regular surveys) was used in the field test of the research project “*sim^{TD}*”. Aim of this study was to compare popular Car-2-X functions (= high willingness to buy, so-called TOP-functions) with unpopular functions (= low willingness to buy, so-called FLOP-functions). Data base of this study is N = 6043 questionnaires and N = 399 drivers of the group-interview-workshops. It is shown that the TOP-functions are evaluated as more helpful, reliable, and trustful than the FLOP-functions. TOP-functions can be characterized by an additional value compared to existing technologies and services. Results are discussed in the context of the *sim^{TD}*-field test.

¹ Interdisziplinäres Zentrum für Verkehrswissenschaften (IZVW), Universität Würzburg, Röntgenring 11, 97070 Würzburg, mvolk@psychologie.uni-wuerzburg.de, tel: +49 931 318 01 35

² Würzburger Institut für Verkehrswissenschaften (WIVW GmbH), Robert-Bosch-Straße 4, 97209 Veitshöchheim, totzke@wivw.de, tel: +49 931 780 09 113

Theoretischer Hintergrund

In den kommenden Jahren steht ein Wandel bezüglich der Ausgestaltung von Informations- und Assistenzsystemen im Fahrzeug bevor: Durch die Einführung von sog. Car-2-X-Technologien stehen den Fahrzeugführern z.B. Informationen zur Verfügung, die sich nicht auf die unmittelbare Umgebung zum Fahrzeug beschränken, sondern für ihn eine deutlich erweiterte Vorausschau erlaubt. Car-2-X-Technologien umfassen z.B. eine Kommunikation zwischen verschiedenen Fahrzeugen, indem im Einzelfahrzeug detektierte Verkehrereignisse (z.B. plötzliche Geschwindigkeitseinbrüche an Stauenden) an nachfolgende und/oder entgegenkommende Verkehrsteilnehmer weitergegeben werden, so dass diese rechtzeitig und angemessen auf die bevorstehende Situation reagieren können. Diese sog. Fahrzeug-Fahrzeug-Kommunikation unterstützt z.B. die Einführung lokaler Gefahrenwarnungen (z.B. Stauende- oder Hinderniswarnungen) aus dem Bereich „Sicherheit“. Zusätzlich wird durch Car-2-X-Technologien die Kommunikation zwischen Fahrzeugen und Verkehrsinfrastruktur (z.B. Lichtsignalanlagen, Verkehrszentralen) möglich, indem fahrzeuginterne Informationen an die Verkehrsinfrastruktur weitergegeben werden, so dass eine Prognose der Verkehrsentwicklung erstellt und den Fahrern eine ideale Fahrtroute zur Verfügung gestellt wird. Hierdurch können z.B. Reisezeiten optimiert oder Staudauern und –längen verringert werden. Zu Funktionen dieser sog. Fahrzeug-Infrastruktur-Kommunikation zählen z.B. die Erweiterte Navigation oder ein Baustelleninformationssystem aus dem Bereich „Verkehr“. Schließlich ermöglichen Funktionen aus dem Bereich „Ergänzende Dienste“ die Weitergabe von zusätzlichen lokalen Informations- und Internetdiensten an die Fahrzeugführer, um diesen z.B. über mögliche Parkmöglichkeiten auf der Strecke zu informieren (z.B. Totzke et al, 2010).

Voraussetzung für die diese verkehrlichen Wirkpotenziale der Car-2-X-Technologien ist die Akzeptanz dieser Technologien durch den Fahrzeugführer: Erst wenn diese Technologien von den Fahrern akzeptiert werden, nutzen die Fahrer diese Technologien und positive Verhaltenseffekte werden sichtbar (z.B. Agarwal & Karahanna, 2000; Fishbein & Ajzen, 1975). Die Erfassung der Nutzerakzeptanz erfolgt dabei regelhaft über Nutzerbefragungen, um die Erwartungen und Einstellungen von Nutzern sowie deren persönliche Erfahrungen zu erfassen (Salcher, 1994). So können z.B. konkrete Systemeigenschaften durch die Nutzer beurteilt und aus der Interaktion mit dem technischen System resultierende Verhaltensänderungen (z.B. Reduktion der Fahrgeschwindigkeit) thematisiert werden. Ebenso sind kaufentscheidende Merkmale des technischen Systems bewertbar (z.B. Grimmer, et al. 1995).

Für eine umfassende Erfassung der Nutzerakzeptanz bietet es sich an, auf eine Kombination aus quantitativen Methoden (z.B. regelmäßige Einzelbefragungen über Fragebögen) und qualitativen Befragungsansätzen (z.B. Gruppeninterviews in Workshop-Form) zurückzugreifen. Gruppeninterviews gelten als motivationsfördernd (Wegge, 2001) und erlauben die Gewinnung von teilweise vorher nicht-

absehbaren Inhalten (Kluck, 2004). Fragebögen bieten demgegenüber die Möglichkeit zu einer stärkeren Standardisierung der Datenerhebung bei einem vertretbaren Personal- und Ressourceneinsatz und lassen dem Probanden einen größeren Freiraum bei der Fragebogenbeantwortung (Raithel, 2006). Fragebogenbefragungen sind zudem in räumlich-zeitlicher Nähe zu relevanten Einzelereignissen durchführbar. Durch einen solchen kombinatorischen Messansatz (bestehend aus qualitativen und quantitativen Methoden sowie Gruppen- und Einzelpersonbezogenen Methoden) wird ein umfassender Blick auf die Nutzerakzeptanz des technischen Systems ermöglicht (z.B. Mayring, 2001): Es wird das Prinzip einer Datentriangulation (Denzin, 1989) realisiert, nach dem mehrere Datenquellen (hier: Befragung von verschiedenen Personen mit unterschiedlichen Methoden) zu Gesamtaussagen integriert und geeignete Schlussfolgerungen gezogen werden.

In der vorliegenden Studie wird auf einen solchen multimethodalen Messansatz zurückgegriffen und die hieraus resultierenden Ergebnisse zur Bewertung von Car-2-X-Technologien durch die Systemnutzer dargestellt. Grundprinzip der Ergebnisdarstellung ist dabei ein Vergleich von bei den Systemnutzern besonders beliebten Car-2-X basierten Funktionen mit eher unbeliebten Funktionen. Es wird herausgearbeitet, wodurch die Beliebtheit einiger Funktionen begründet zu sein scheint. Das Ausmaß der Kaufbereitschaft dient dabei als Gruppierungsmerkmal, um die bei Versuchsfahrern beliebte von unbeliebten Funktionen abzugrenzen.

Diese Studie ist Teil des Forschungsprojekts „Sichere Intelligente Mobilität – Testfeld Deutschland“ (sim^{TD}), das eine empirische Erprobung der Car-2-X-Technologien sowohl im Realverkehr (Feldversuch und Testgelände) als auch in einem Simulationslabor (bestehend aus Fahr- und Verkehrssimulation) zum Ziel hatte. Es wird nachfolgend ausschließlich auf Ergebnisse des Feldversuchs eingegangen. Das Projekt wurde durch das Bundesministerium für Wirtschaft und Technologie, das Bundesministerium für Bildung und Forschung sowie das Bundesministerium für Verkehr, Bau und Straßenentwicklung gefördert.

Methoden

Vorgehen während des Feldversuchs

Im Rahmen des sim^{TD}-Feldversuchs wurden 22 Funktionen eingesetzt, die auf Car-2-X-Technologien basieren. Solche sind beispielsweise für die o.g. Bereiche:

- Bereich „Verkehr“: Verkehrs(fluss)informationen und –steuerung (z.B. Straßenvorausschau, Umleitungsmanagement, Ampelphasenassistent: Grüne Welle und Restrotanzeige)
- Bereich „Sicherheit“: Lokale Gefahrenwarnungen und Fahrerassistenz: (z.B. Hindernis-, Stauendewarnung, Ampelphasenassistent: Rotlichtverstoß)
- Bereich „Ergänzende Dienste“: Lokale Informationsdienste (z.B. Standortinformationsdienste)

Diese Funktionen wurden im Zeitraum von Juli bis Dezember 2012 täglich mit 120 Fahrzeugen und drei Motorrädern in Frankfurt/Main und Umgebung untersucht. Als Versuchsgebiet fungierten öffentliche Straßen des Rhein-Main Gebiets (Innenstadt, Landstraße und Autobahn) sowie ein Testgelände in Friedberg/Hessen.

Die Versuchsfahrer wurden täglich in 30er-Gruppen, welche denselben Tagesablauf hatten, instruiert und durchführen vorgegebene Versuchsszenarien (sog. Drehbücher), welche eine empirische Überprüfung der Car-2-X-Technologien unter variierenden Rahmenbedingungen (z.B. Verkehrsdichte oder Streckenart) ermöglichten. Ein Versuchstag (bestehend aus Instruktion, Organisation, Fahrten und Befragungen) dauerte pro Fahrergruppe ca. fünf Stunden. Die Versuche wurden regelmäßig im Sinne eines Kontrollgruppen-Designs durchgeführt (d.h. ein Teil der Versuchsfahrer erhielt Meldungen über ein eigenständiges Display in der oberen Mittelkonsole des Versuchsfahrzeugs, während einem anderen Teil der Versuchsfahrer keine Meldungen präsentiert wurden). Die Fahrer stammten aus einem Fahrerpool, der für diesen Feldversuch aufgebaut wurde und nach jeweils acht Wochen komplett ausgetauscht wurde. Hieraus unterteilte sich der gesamte Versuchszeitraum in drei Versuchsblöcke.

Erfassung der Nutzerakzeptanz

Es kamen folgende Messansätze zur Erfassung der Nutzerakzeptanz zum Einsatz:

1. Gruppeninterviews in Workshop-Form vs.
2. Einzelbefragungen mittels Fragebögen unmittelbar nach Versuchsfahrten

Im Rahmen der Workshops wurden in Form von Gruppeninterviews und Kleingruppenarbeiten Beurteilungen der sim^{TD}-Funktionen vorgenommen. Die Workshops wurden jeweils am Blockanfang und –ende durchgeführt. Nachfolgend wird ausschließlich auf Ergebnisse der Workshops am Blockende eingegangen. In diesen Workshops erhielten die Versuchsfahrer u.a. eine Auflistung aller in sim^{TD}-erprobten Funktionen. Die Fahrer sollten zunächst angeben, ob sie diese bereits erlebt hatten („Bin ich schon gefahren“, Ja-Nein-Antwortformat). Sofern Sie diese Frage bejahten, wurde auch die Kaufbereitschaft bezüglich der erprobten Funktionen („Ich würde diese Funktion kaufen.“; Ja-Nein-Antwortformat) sowie die Zuverlässigkeit der jeweiligen Funktion („Die Meldung erscheint in der Regel zuverlässig.“; 7-stufige Skala mit Ankern „trifft nicht zu“ und „trifft zu“) abgefragt.

Des Weiteren wurden regelmäßig Einzelbefragungen in Form von Fragebögen unmittelbar nach der jeweiligen Versuchsfahrt durchgeführt. Der Schwerpunkt lag in der Bewertung der Fahrererfahrungen in Interaktion mit den sim^{TD}-Funktionen während der vorherigen Fahrt. Im Folgenden wird auf diese Items (7-stufige Skala mit Ankern „lehne stark ab“ und „stimme stark zu“) eingegangen:

- „Die Meldung/Anzeige war hilfreich.“

- „Ich habe der Meldung vertraut.“ (für Funktionen aus Bereich „Sicherheit“ und „Verkehr“)
- „Die Warnung hat mir geholfen, die Gefahr früher wahrzunehmen.“ (für Funktionen aus Bereich „Sicherheit“)

Zusätzlich wurden Stärken und Schwächen der Funktionen bewertet:

- „Was finden Sie gut an [Funktionsname]?“ (offenes Antwortformat)
- „Was finden Sie schlecht an [Funktionsname]?“ (offenes Antwortformat)

Nachfolgend werden Funktionen, für die die Versuchsfahrer die insgesamt höchste Kaufbereitschaft äußerten, solchen Funktionen mit der niedrigsten Kaufbereitschaft gegenübergestellt. Die erstgenannten Funktionen werden als „TOP-Funktionen“, die letztgenannten Funktionen als „FLOP-Funktionen“ bezeichnet. Sofern nicht anders erwähnt, wird eine einfaktorielle Varianzanalyse mit dem Faktor „Klassifikation“ (TOP vs. FLOP) verwendet.

Beschreibung der Stichprobe

Grundlage der Ergebnisdarstellungen sind N = 6043 Einzelbefragungs-Fragebögen von N = 341 Versuchsfahrern (n = 260 Männer, n = 81 Frauen) sowie N = 399 Probanden (n = 314 Männer, n = 85), die an den Gruppeninterview-Workshops teilgenommen haben. Tab. 1 listet ausgewählte Deskriptiva der entsprechenden Stichproben. An dieser Stelle werden Daten aus den Versuchsblöcken 2 und 3 berücksichtigt. Daten aus Versuchsblock 1 werden nicht weiter berücksichtigt, da dieser Versuchsblock im Sinne eines Vorversuchs zu verstehen ist.

Tab. 1: Stichprobenbeschreibungen für die Versuchsblöcke 2 und 3.

Versuchsblock	Methode	Stichprobe [Anzahl] Gesamt (Männer, Frauen)	Alter [in Jahren] M (SD)
2 (26.08.-19.10.2012)	Fragebogen	172 (129 m, 43 w)	33.33 (11.81)
	Workshop	185 (141 m, 44 w)	33.62 (12.11)
3 (22.10.-14.12.2012)	Fragebogen	169 (131 m, 38 w)	33.86 (11.64)
	Workshop	214 (173 m, 41 w)	34.68 (12.26)

Die Gruppeninterview-Workshops fanden jeweils in den ersten beiden Wochen (Blockbeginn) bzw. in der letzten Woche eines Versuchsblocks (Blockende) statt. Es wurden zu Blockbeginn jeweils fünf Workshops bzw. zu Blockende jeweils vier Workshops durchgeführt. Ein Workshop dauerte ca. 2.5 Std., an welchem jeweils zwischen 16 und 31 Probanden teilnahmen. Die Fragebögen wurden im Verlauf der beiden Versuchsblöcke ausgegeben.

Ergebnisse

Abb. 1 veranschaulicht die sim^{TD} -Funktionen mit der höchsten Kaufbereitschaft (sog. TOP-Funktionen) sowie die Funktionen mit der geringsten Kaufbereitschaft (sog. FLOP-Funktionen). Zu den TOP-Funktionen zählen: Stauendewarnung, Längsführungsassistent, Hinderniswarnung (alle aus Bereich „Sicherheit“), Ampelphasenassistent (Restrotanzeige und Grüne Welle; Bereich „Verkehr“) sowie Standortdienste (Parksituation; Bereich „Ergänzende Dienste“). Als FLOP-Funktionen haben sich im sim^{TD} -Feldversuch herausgestellt: Straßenwetterwarnung, Ampelphasenassistent (Rotlichtverstoß) und Verkehrszeichen-Assistent (Bereich „Sicherheit“), Straßenvorausschau (Bereich „Verkehr“) sowie Standortdienste (Kommunal- und Wetterinformation; Bereich „Ergänzende Dienste“).

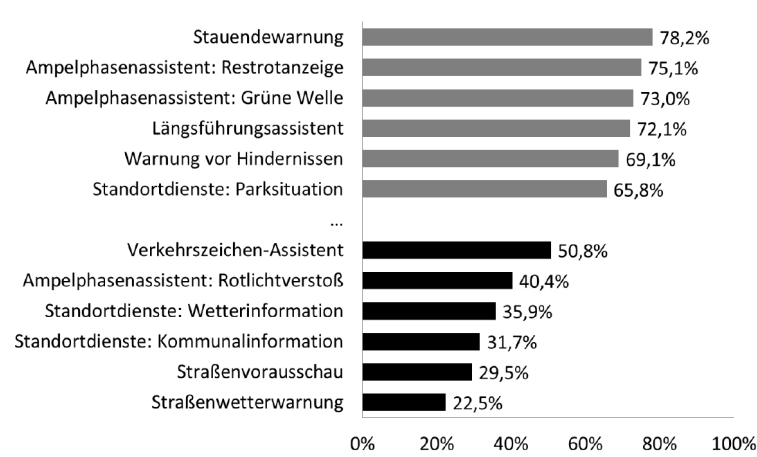


Abb. 1.: Prozentuale Häufigkeit der Zustimmung auf Frage „Ich würde diese Funktion kaufen“. Dargestellt sind Funktionen mit höchster Kaufbereitschaft (TOP-Funktionen, oben) bzw. mit niedrigster Kaufbereitschaft (FLOP-Funktionen, unten).

Hinweise für Unterschiede in der Bewertung der TOP- und FLOP-Funktionen durch die Fahrer ergeben sich aus den Ergebnissen der Einzelbefragungen: So werden TOP-Funktionen generell als hilfreicher eingestuft als FLOP-Funktionen (Funktionen des Bereichs „Sicherheit“: $F_{1,663} = 29.72$, $p = .000$, $\eta^2 = .04$; Bereich „Verkehr“: $F_{1,609} = 6.74$, $p = .010$, $\eta^2 = .01$; Bereich „Ergänzende Dienste“: $F_{1,785} = 4.10$, $p = .043$, $\eta^2 = .01$; siehe Abb. 2). TOP-Funktionen aus dem Bereich „Sicherheit“ helfen zudem, Gefahren auf der Straße früher wahrzunehmen ($F_{1,661} = 16.28$, $p = .000$, $\eta^2 = .02$; siehe Abb. 2 Mitte). Die TOP-Funktionen werden als zuverlässiger beurteilt (über alle Bereiche hinweg: $F_{1,3397} = 49.31$, $p = .000$, $\eta^2 = .03$), so dass den TOP-Funktionen stärker vertraut wird („Sicherheit“: $F_{1,658} = 49.41$, $p = .000$, $\eta^2 = .07$; „Verkehr“: $F_{1,483} = 12.08$, $p = .001$, $\eta^2 = .02$; siehe Abb. 2 rechts).

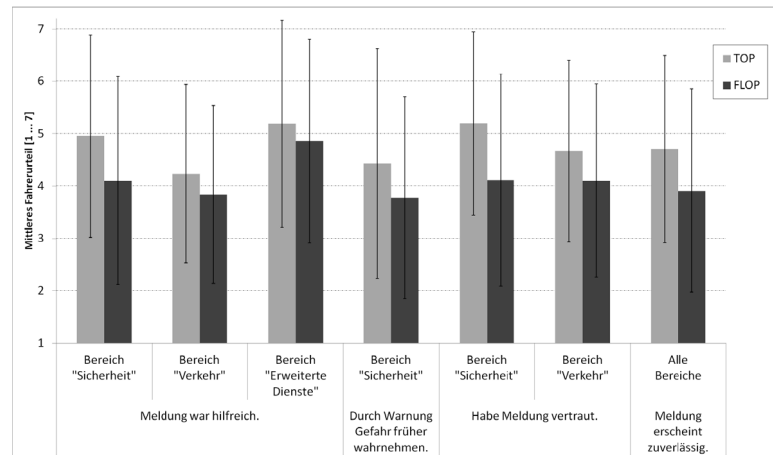


Abb. 2.: Fahrerurteile für ausgewählte Fragen für TOP- und FLOP-Funktionen. Dargestellt sind Mittelwerte mit Standardabweichung.

Weitere Begründungen für die Präferenz der TOP-Funktionen resultieren aus den freien Kommentaren der Versuchsfahrer in den Fragebögen. Am häufigsten werden als Stärke der TOP-Funktionen genannt (Angaben in Klammern: prozentuale Anteil an Gesamtzahl der geäußerten Argumente):

- Bereich „Sicherheit“: Sicherheitsgewinn (62.4%) aufgrund frühzeitiger Warnung (16.9%)
- Bereich „Verkehr“: Effizienzgewinn (44.4%) aufgrund eines Mehrwerts der dargestellten Informationsinhalten (z.B. die optimale Geschwindigkeitsanzeige bzw. restliche Dauer der Rotphase, 30.2%)
- Bereich: „Ergänzende Dienste“: Effizienzgewinne (z.B. Zeitersparnis oder optimaler Verkehrsfluss, 20.8%) und ihre Informationsinhalte (z.B. Informationen über die Art der Parkplätze oder Parkgebühren, 53.9%)

Diskussion und Ausblick

Ziel der vorliegenden Studie war es, Hinweise zu geben, welche Car-2-X-Technologien von Fahrzeugführern besonders stark akzeptiert werden sowie mögliche Hinweise für diese Präferenzen abzuleiten. Hierzu wurden Fahrer mittels verschiedener Methoden (Gruppeninterviews in Workshops vs. regelmäßige Einzelbefragungen mittels Fragebögen) befragt. Bei den Fahrern beliebte Car-2-X basierte Funktionen wurden mit eher unbeliebten Funktionen verglichen. Das Spektrum möglicher Funktionen umfasste die Bereiche „Verkehr“, „Sicherheit“ und „Ergänzende Dienste“, wie sie im Forschungsprojekt sim^{TD} zum Einsatz kamen.

Im Rahmen des sim^{TD} -Feldversuchs wird für folgende Car-2-X Funktionen die höchste Kaufbereitschaft geäußert: Stauendewarnung, Längsführungsassistent, Hinderniswarnung, Ampelphasenassistent (Restrotanzeige und Grüne Welle) sowie Standortdienste (Parksituation). Diese bei den Fahrern beliebten Funktionen werden zugleich als hilfreicher, zuverlässiger und vertrauenswürdiger bewertet als Funktionen, die hinsichtlich der Kaufbereitschaft nicht so positiv abschneiden. Ins-

besondere der wahrnehmbare Mehrwert der Funktionen im Vergleich zu bestehenden Technologien beziehungsweise Services ist dabei von Bedeutung: So zeichnen sich beliebte Funktionen u.a. durch die Darstellung zusätzlicher Informationsinhalte (z.B. einer Anzeige der optimalen Fahrgeschwindigkeit zum Durchfahren eines Streckenabschnitts) aus, die zurzeit noch nicht in den Fahrzeugen zur Verfügung stehen.

Das methodische Vorgehen dieser Studie kann zusammenfassend positiv bewertet werden: Durch die Verschränkung qualitativer und quantitativer Methoden wurden verschiedenartige Messansätze zur Erfassung der Nutzerakzeptanz umgesetzt, wodurch eine umfassende Betrachtung der Nutzerakzeptanz von Car-2-X-Technologien ermöglicht wurde. Nichtsdestotrotz sind bei einer Bewertung der vorliegenden Befunde die gegebenen Rahmenbedingungen zu beachten: So wurde im Rahmen des Projekts *sim^{TD}* kein serienreifes Car-2-X System eingesetzt, sondern ein Forschungssystem entwickelt und empirisch erprobt. Hierdurch wurden für die Versuchsfahrer möglicherweise Systemfehler und –grenzen sichtbar, was einen negativen Einfluss auf die Fahrerurteile haben kann. Darüber hinaus wurden im Rahmen des Feldversuchs instruierte Fahrten realisiert (d.h. vorgegebene Routen wurden in Form sog. Drehbücher abgefahren). Die Fahrer hatten zudem über die Dauer von jeweils 8 Wochen täglich bis zu fünf Stunden Kontakt mit dem Car-2-X System. Die Erfahrungen der Versuchsfahrer sind daher nicht ohne weiteres mit einem alltäglichen Systemkontakt zu vergleichen. Um alltagsnähere Ergebnisse zu ermöglichen, wären daher gegebenenfalls Field Operational Tests zu fordern, in denen Versuchsfahrer frei mit Car-2-X-Technologien interagieren können.

Literatur

- Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage 1. *MIS quarterly*, 24 (4), 665 – 694.
- Denzin, N.K. (1989). *The Research Art: A Theoretical Introduction to Sociological Methods* (3rd ed.). Chicago: Aldine.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Grimmer, W., Adelt, P.J., & Stephan, E.R. (1995). *Die Akzeptanz von Navigations- und Verkehrsführungssystemen der Zukunft: Eine AXA-Direkt-Verkehrsstudie*. Bonn: Deutscher Psychologen Verlag.
- Kluck, M. (2004). Methoden der Informationsanalyse – Einführung in die empirischen Methoden für die Informationsbedarfsanalyse und die Markt- und Benutzerforschung. In R. Kuhlen, T. Seeger, & D. Strauch (Hrsg.), *Grundlagen der praktischen Information und Dokumentation. Handbuch zur Einführung in die Informationswissenschaft und –praxis* (5. Ausg.), Bd. 1 (S. 271 – 288). München: K.G. Saur.

- Mayring, P. (2001). Combination and integration of qualitative and quantitative analysis. In *Forum Qualitative Sozialforschung / Forum Qualitative Social Research*, 2 (1). Available at: <http://www.qualitative-research.net/fqs/fqs.htm> (Date of access: 05-MAY-2012).
- Raithel, J. (2006). *Quantitative Forschung: Ein Praxiskurs*. Wiesbaden: VS-Verlag.
- Salcher, E.F. (1995). *Psychologische Marktforschung*. Berlin: De Gruyter.
- Totzke, I., Mühlbacher, D., Buld, S., & Krüger, H.-P. (2010). sim^{TD} (Sichere Intelligente Mobilität - Testfeld Deutschland): Empirische Wirkungsermittlung von Car-2-X-Technologien im Feldversuch und in der Fahrsimulation. In *Ta-gungsband 19. Aachener Kolloquium Fahrzeug- und Motorentechnik* (S. 1701 – 1712). Aachen: fka Forschungsgesellschaft Kraftfahrwesen mbH Aachen.
- Wegge, J. (2001). Gruppenarbeit. In H. Schuler (Hrsg.), *Lehrbuch der Personal-psychologie* (S.484 – 507). Göttingen: Hogrefe.

Modulare Evaluation interaktiver Technik. Entwicklung und Validierung des meCUE Fragebogens zur Messung der User Experience.

Michael Minge¹, Laura Riedel¹ und Manfred Thüning²

Schlüsselwörter: User Experience, Usability, Emotionen, Evaluation, Fragebogen.

Zusammenfassung

Während der menschenzentrierte Gestaltungsprozess noch vor kurzer Zeit vor allem darauf abzielte, eine möglichst hohe Gebrauchstauglichkeit sicherzustellen, geht es mittlerweile darum, auch die Wahrnehmung und das Erleben des Nutzers bei Interaktion mit technischen Geräten zu berücksichtigen und positiv zu beeinflussen. Zur adäquaten Erfassung dieses vielschichtigen Konzepts „User Experience“ wird ein theoriebasierter Fragebogen zur modularen Evaluation von Technik (*meCUE*) vorgestellt. Dieser Beitrag fasst die bisherigen Entwicklungsschritte bei der Konstruktion und Validierung des Fragebogens zusammen. Speziell wird die empirische Überprüfung einer Minimalrevision und einer inhaltlichen Erweiterung behandelt.

Theoretischer Hintergrund

Die Berücksichtigung des Nutzungserlebens („*User Experience*“) stellt Entwickler und Hersteller von Technik vor die Herausforderung, subjektive Bewertungen aus Nutzerperspektive in einer geeigneten Art und Weise zu erheben und die erzielten Ergebnisse im menschenzentrierten Gestaltungsprozess nachhaltig zu berücksichtigen (DIN EN ISO 9241-2010, 2011). Neben qualitativen Verfahren zur systematischen Ermittlung persönlicher Eindrücke eröffnen Fragebögen die Möglichkeit zur standardisierten und ökonomischen Erfassung subjektiver Bewertungen.

Um relevante Aspekte im Nutzungserleben möglichst ganzheitlich zu erfassen, wurde auf Basis eines weithin etablierten Modells, dem CUE-Modell von Thüning und Mahlke (2007), ein modular aufgebauter Fragebogen entwickelt. In Anlehnung an Hassenzahl (2001) unterscheidet das CUE-Modell zwischen der Wahrnehmung aufgabenbezogener und nicht-aufgabenbezogener Produktqualitäten und berücksichtigt darüber hinaus Nutzeremotionen als einen mediiierenden Faktor für die Ausformung bestimmter Konsequenzen, wie z.B. der Akzeptanz (siehe Abb. 1).

Die Zielstruktur des neuen Fragebogens wurde darauf aufbauend entwickelt. Als zentrale Aspekte wurden dabei die separat anwendbaren Module „Produktwahrnehmungen“ (Nützlichkeit, Benutzbarkeit, visuelle Ästhetik, Status, Bindung), „Nutzeremotionen“ (positive und negative Emotionen) und „Konsequenzen“ (Produktloyalität, Nutzungsintention) mit den genannten Subskalen definiert.

¹ Technische Universität Berlin, Graduiertenkolleg prometei

² Technische Universität Berlin, Fachgebiet Kognitionspsychologie und Kognitive Ergonomie

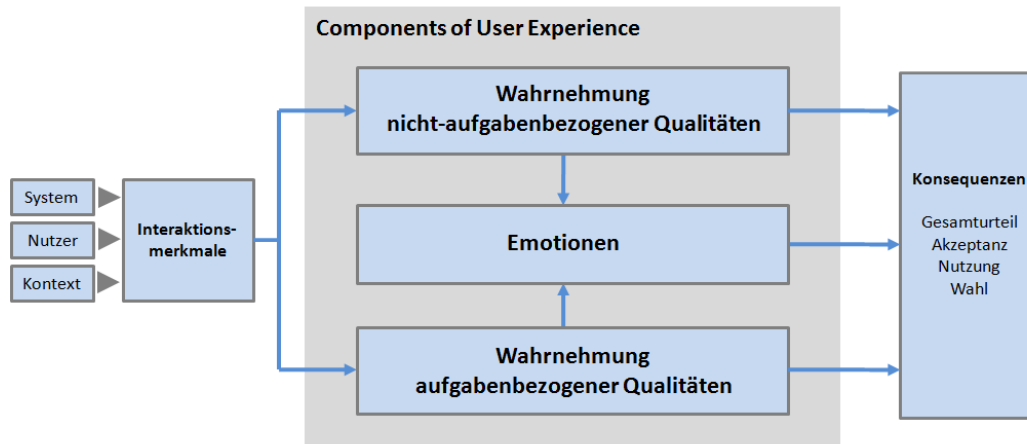


Abb. 1: Modell des Nutzungserlebens (CUE) nach Thüring und Mahlke (2007, S. 262)

Empirische Vorarbeiten

Die Konstruktion des Fragebogens zur modularen Evaluation zentraler Aspekte der User Experience (*meCUE*) erfolgte auf Basis zweier Online-Datenerhebungen. Der anfängliche Itempool bestand aus 67 Items, die deduktiv entsprechend der Zielstruktur formuliert und den Subskalen theoriegeleitet zugeordnet wurden. Alle Items waren als Aussagen formuliert, deren Zustimmung über ein Likert skaliertes Antwortformat erfasst wird („lehne völlig ab“, „lehne ab“, „lehne eher ab“, „weder noch“, „stimme eher zu“, „stimme zu“, „stimme völlig zu“). Beide Stichproben (jeweils $n = 238$ Probanden) beinhalteten Bewertungen zu technischen Produkten unterschiedlicher Art. Die Itemselektion und Validierung erfolgte durch separate Faktorenanalysen. Im Ergebnis entstand ein Fragebogen, der die Zielstruktur über 33 Items erfasst und solide Werte sowohl hinsichtlich der Varianzaufklärung in den Modulen „Produktwahrnehmungen“ (69,6 %), „Emotionen“ (57,4 %) und „Konsequenzen“ (63,5 %) aufwies als auch hinsichtlich der internen Konsistenz seiner Subskalen mit *Cronbachs Alpha* Werten zwischen .69 und .83 (Minge & Riedel, accepted).

Die faktorielle Struktur konnte mit den Daten aus einer laborexperimentellen Untersuchung stabil repliziert werden (Minge & Riedel, accepted). Hierbei bewerteten $n = 67$ Probanden jeweils drei interaktive Produkte aus verschiedenen Bereichen (Textbearbeitungssoftware, Audio-Player und Mobiltelefon). Aufgrund größerer Varianzen zwischen den in diesem Fall vorgegebenen Produkten zeigten sich die Anteile erklärter Varianz bei Analyse der vorliegenden 201 Datensätze nochmals erhöht (81,1 % für „Produktwahrnehmungen“, 74,3 % für „Emotionen“ und 74,1 % für „Konsequenzen“). Ebenso unterstützten die Daten erneut die Annahme guter interner Konsistenz aller Subskalen ($.83 < \text{Cronbachs Alpha} < .94$).

Um die Validität des Verfahrens abzuschätzen, wurden im Rahmen der Laborstudie Zusammenhänge zu bestehenden Fragebögen erhoben. Hierbei zeigten sich substantiell hohe Korrelationen vor allem zwischen den Dimensionen „Benutzbarkeit“ (*meCUE*), „pragmatische Qualität“ (*AttrakDiff* nach Hassenzahl, Burmester

& Koller, 2008) sowie „Durchschaubarkeit“ und „Vorhersagbarkeit“ (*UEQ* nach Laugwitz, Schrepp & Held, 2006). Werte der „visuellen Ästhetik“ (*meCUE*) waren hoch verbunden mit den Dimensionen klassischer und expressiver Ästhetik nach Lavie und Tractinsky (2004). Im Bereich Emotionen lagen signifikante Korrelationen zwischen den Dimensionen des *meCUE* und den Skalen des *PANAS* (Watson, Clark & Tellegen, 1988) vor. Erwartungsgemäß korrelierte das Außenkriterium der bearbeiteten Aufgabenmenge mit den subjektiven Angaben zur „Nützlichkeit“ und „Benutzbarkeit“ (*meCUE*), wohingegen zu den nicht-aufgabenbezogenen Dimensionen keine substanziellen Korrelationen vorlagen (Minge & Riedel, accepted).

Zusammenfassend zeigen diese Arbeiten auf, dass die theoriegeleitete Konstruktion eines zuverlässigen und validen Messinstrumentes zur Erhebung zentraler Aspekte des Nutzungserlebens erfolgreich gelungen ist. Hinsichtlich der Formulierung einzelner Items und der angebotenen Module bestand allerdings noch Handlungsbedarf. Dieser wurde durch eine Minimalrevision und eine inhaltliche Erweiterung des *meCUE* Fragebogens abgedeckt und mit der folgenden Studie überprüft.

Zielsetzung

Das Ziel der Studie bestand einerseits darin, den revidierten Fragebogen hinsichtlich seiner faktoriellen Struktur zu überprüfen und andererseits die Validität eines neuen Moduls zur Erfassung eines Gesamturteils zu bewerten. Die Revision umfasste die Abwandlung zweier Items, wobei ein Item umformuliert wurde, um eine einheitliche Polung in der Subskala „Benutzbarkeit“ zu erzielen und ein weiteres Item, um dessen inhaltliche Aussagekraft zu fokussieren (siehe Tab. 1).

Tab. 1: umformulierte Items in der minimal revidierten Version

bisherige Formulierung	revidierte Formulierung
Es dauert zu lange, die Funktionen des Produkts zu erleben.	Es wird schnell klar, wie das Produkt zu bedienen ist.
Insgesamt halte ich das Produkt für nützlich.	Ich halte das Produkt für absolut nützlich.

Erweitert wurde die modulare Struktur des Fragebogens um die Erhebung eines Gesamturteils (Modul 4). Verfügbare Fragebögen zum Nutzungserleben (z.B. *AttrakDiff* und *UEQ*) erfassen dieses Gesamturteil unter der Bezeichnung „Attraktivität“ als zusätzliche Subskala. In *meCUE* wurde die Skala als feinabgestuftes semantisches Differenzial mit den Polen „schlecht“ und „gut“ realisiert. Die bisherige Umsetzung im Fragebogentool LimeSurvey® gestattet, die Auswahl zwischen den insgesamt 21 Abstufungen über einen Schieberegler vorzunehmen (siehe Abb. 2). Die durch dieses Modul erweiterte Struktur von *meCUE* zeigt Abbildung 3.



Abb. 2: Realisierung des Moduls 4 als Schieberegler zur Erhebung des Gesamturteils

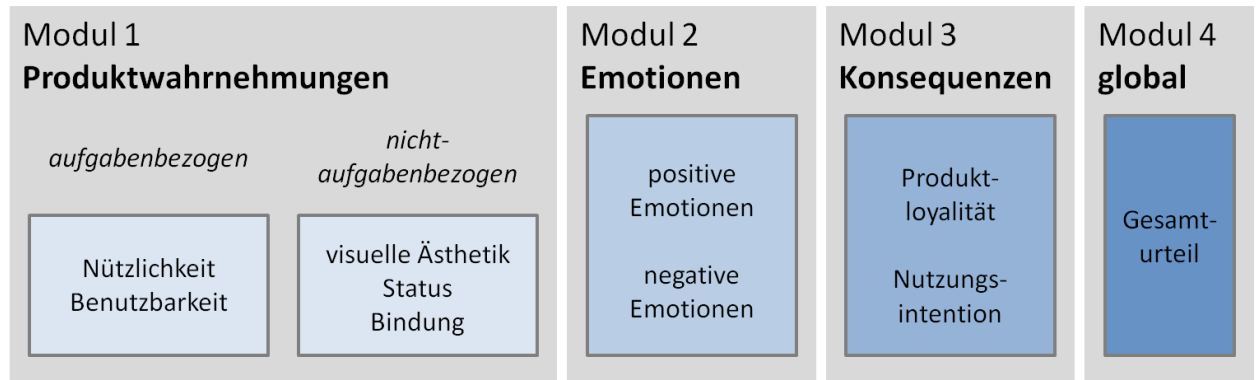


Abb. 3: modulare Struktur der revidierten Version des meCUE Fragebogens

Methode

Zur Überprüfung der faktoriellen Struktur des überarbeiteten Fragebogens wurde eine Online-Datenerhebung durchgeführt. Hierzu waren $n = 237$ Probanden aufgefordert, jeweils ein interaktives technisches Produkt aus ihrem Alltag anhand des minimal revidierten *meCUE* Fragebogens zu bewerten. Die Stichprobe setzte sich aus 139 Teilnehmerinnen und 98 Teilnehmern zusammen und wies ein Durchschnittsalter von 29,8 Jahren auf. Die vier Module des Fragebogens wurden in der in Abbildung 3 beschriebenen Reihenfolge fest vorgegeben, während die einzelnen Items innerhalb eines jeden Moduls randomisiert präsentiert wurden.

Um die Validität des neu konstruierten Schiebereglers zur Erhebung des globalen Produkturteils (Modul 4) abzuschätzen, wurde zusätzlich zum *meCUE* auch der *AttrakDiff* Fragebogen von Hassenzahl, Burmester und Koller (2008) eingesetzt. Entsprechend der theoretischen Vorüberlegungen wurden hierbei substanzielle Zusammenhänge zwischen dem Modul 4 des *meCUE* und der Dimension „Attraktivität“ des *AttrakDiff* erwartet. Wie sich der Datensatz hinsichtlich Art und Häufigkeit der bewerteten technischen Produkte zusammensetzt, ist in Tabelle 2 aufgeführt.

Tab. 2: Art und Häufigkeit der bewerteten Produkte in der Stichprobe

Produktbereich	absolute Häufigkeit	relative Häufigkeit
Mobiltelefon	99	41,8 %
Laptop/ Computer	75	31,6 %
Kaffeemaschine	22	9,3 %
mobiler Audioplayer (MP3)	16	6,8 %
Smartphone App	9	3,8 %
Waschmaschine	7	3,0 %
Digitalkamera	6	2,5 %
Software	3	1,3 %
Summe	237	100,0 %

Ergebnisse

In einem ersten Schritt erfolgte die modulweise Validierung der revidierten Fragebogenversion durch separate Faktorenanalysen der mehrdimensionalen Module.

Auf Basis des Minimum-Average-Partial-Tests (MAP-Test nach Velicer, 1976) ergab die varianzmaximierende Hauptkomponentenanalyse der zugrundeliegenden 15 Items des Moduls „Produktwahrnehmungen“ die erwarteten fünf Hauptkomponenten. Insgesamt erklären diese fünf Faktoren 73,7 % der aufgetretenen Varianz (siehe Tab. 3). Die ebenfalls in Tabelle 3 dargestellte faktorielle Struktur zeigt, dass die bei der Konstruktion etablierte Struktur stabil repliziert wird und die inhaltliche Bezeichnung der Subskalen weiterhin aufrecht erhalten werden kann. Auf Basis der *Cronbachs Alpha* Koeffizienten ist die interne Konsistenz der Subskalen als ausreichend bis gut zu bezeichnen.

Tab. 3: Faktorladungen, Cronbachs Alpha und Varianzaufklärung der Subskalen im Modul „Produktwahrnehmung“ (dokumentiert sind alle Ladungen > .4)

Item:	Nützlich- keit	Benutz- barkeit	visuelle Ästhetik	Status	Bindung
Mithilfe des Produkts kann ich meine Ziele erreichen.	.836				
Ich halte das Produkt für absolut nützlich.	.765				
Die Funktionen des Produkts sind genau richtig für meine Ziele.	.654				
Es wird schnell klar, wie das Produkt zu bedienen ist.		.863			
Die Bedienung des Produkts ist verständlich.		.859			
Das Produkt lässt sich einfach benutzen.		.853			
Das Design wirkt attraktiv.			.876		
Das Produkt ist kreativ gestaltet.			.845		
Das Produkt ist stilvoll.			.828		

Durch das Produkt werde ich anders wahrgenommen.				.875	
Das Produkt verleiht mir ein höheres Ansehen.				.830	
Meine Freunde dürfen ruhig neidisch auf das Produkt sein.			.401	.578	
Ohne das Produkt kann ich nicht leben.					.827
Wenn ich das Produkt verlieren würde, würde für mich eine Welt zusammenbrechen.					.814
Das Produkt ist wie ein Freund für mich.					.756
Cronbachs Alpha	.73	.86	.86	.76	.80
Varianzaufklärung	12,9	16,3	16,0	13,2	15,3

Im Modul „Nutzeremotionen“ ergab die varianzmaximierende Faktorenanalyse der zwölf Items ebenfalls die erwarteten zwei zugrundeliegenden Faktoren für das Erleben positiver und negativer Emotionen. Gemeinsam erklären diese 66,6 % der aufgetretenen Varianz. Zudem liegen für beide Subskalen Cronbachs Alpha-Werte vor, die die Annahme guter interner Konsistenz unterstützen (siehe Tab. 4).

Tab. 4 : Faktorladungen, Cronbachs Alpha und Varianzaufklärung der Subskalen im Modul „Nutzeremotionen“ (dokumentiert sind alle Ladungen > .4)

Item:	positive Emotionen	negative Emotionen
Das Produkt entspannt mich.	.858	
Durch das Produkt fühle ich mich ausgeglichen.	.846	
Durch das Produkt fühle ich mich fröhlich.	.824	
Das Produkt beruhigt mich.	.822	
Das Produkt beschwingt mich.	.809	
Das Produkt stimmt mich euphorisch.	.805	
Das Produkt frustriert mich.		.866
Das Produkt nervt mich.		.826
Das Produkt macht mich müde.		.818
Das Produkt verärgert mich.		.800
Durch das Produkt fühle ich mich erschöpft.		.781
Durch das Produkt fühle ich mich passiv.		.625
Cronbachs Alpha	.91	.88
Varianzaufklärung	35,0	31,6

Bei Analyse der verbliebenen sechs Items im Modul „Konsequenzen“ wurden unter Bezugnahme auf den MAP-Test zwei Faktoren extrahiert, die zusammen 65,5 % der aufgetretenen Varianz erklären. Die Verteilung der Ladungsmuster auf beide Dimensionen unterstützt die nach Abschluss der Konstruktionsphase formu-

lierte inhaltliche Benennung. Die Cronbachs Alpha-Werte sprechen für eine ausreichende interne Konsistenz der zwei Subskalen (siehe Tab.5).

Tab. 5: Faktorladungen, Cronbachs Alpha und Varianzaufklärung der Subskalen im Modul „Konsequenzen“ (dokumentiert sind alle Ladungen > .4)

Item:	Produkt-loyalität	Nutzungs-intention
Ich würde das Produkt gegen kein anderes eintauschen.	.873	
Im Vergleich zu diesem Produkt wirken andere Produkte unvollkommen.	.855	
Ich würde mir genau dieses Produkt jederzeit (wieder) zulegen.	.597	.494
Ich kann es kaum erwarten, das Produkt erneut zu verwenden.		.840
Wenn ich könnte, würde ich das Produkt täglich nutzen.		.755
Wenn ich mit dem Produkt zu tun habe, vergesse ich schon mal die Zeit.		.711
Cronbachs Alpha	.70	.75
Varianzaufklärung	32,3	33,2

In einem zweiten Schritt wurden die Angaben zum globalen Produkturteil ausgewertet und auf Zusammenhänge untersucht. Einerseits wurde das Gesamturteil über einen Schieberegler im neu konstruierten Modul 4 des *meCUE* erhoben, andererseits über die *AttrakDiff* Dimension „Attraktivität“, die wiederum das arithmetische Mittel aus Antworten auf insgesamt sieben Items darstellt. Die substantielle Korrelation in Höhe von $r = .559$ zwischen beiden Variablen unterstützt die Annahme konvergenter Validität des Schiebereglers. Statistisch bedeutsame Korrelationen, allerdings von geringerer Höhe, liegen darüber hinaus zu den übrigen Subskalen des *AttrakDiff* vor (siehe Tab. 6).

Tab. 6: Korrelationen zwischen dem Modul 4 (meCUE) und den AttrakDiff Dimensionen
Anmerkung: * $p < .05$, ** $p < .01$

Korrelation r	<i>AttrakDiff</i>			
	Pragmatik	Stimulation	Identifikation	Attraktivität
<i>meCUE</i> Gesamturteil	.477**	.448**	.439**	.559**

Diskussion

Die Ergebnisse aus der vorliegenden Studie unterstützen die bisherigen Befunde zur Qualität und Güte des theoriebasiert entwickelten Fragebogens zur modularen Evaluation zentraler Aspekte des Nutzungserlebens, *meCUE*. Insbesondere zeichnet sich der *meCUE* Fragebogen durch eine besonders stabile faktorielle Struktur und eine ausreichende bis gute interne Konsistenz seiner Subskalen aus. Mit der Revision wurde ein konsistentes Antwortformat realisiert, mit dem sich Nutzerbewertungen in kurzer Zeit (ca. 3-5 Minuten) äußerst effizient einholen lassen.

Die Erweiterung des Fragebogens um ein viertes Modul zur Erhebung eines allgemeinen Gesamturteils kann ebenfalls als gelungen bezeichnet werden. Der er-

mittelte Zusammenhang zu vergleichbaren Skalen ist, insbesondere vor dem Hintergrund der unterschiedlichen Skalierung beider Variablen und der Tatsache, dass es sich bei dem *meCUE* Gesamturteil um eine Single-Item Lösung handelt, substantiell und ausreichend hoch. Für den Fall, dass verschiedene Gestaltungslösungen miteinander verglichen werden sollen, erlaubt die Erhebung eines Gesamturteils aus anwendungsorientierter Sicht bei inferenzstatistischer Überprüfung vorliegender Unterschiede das Aufstellen einer eindeutigen Rangfolge der Untersuchungsobjekte. Aus forschungstheoretischer Sicht eröffnet sich durch die Erhebung des Gesamturteils die Möglichkeit zur Bildung eines Regressionsmodells, um die allgemeine Produktgüte vorherzusagen. Dadurch werden zukünftig nicht nur zusätzliche Aussagen zur strukturellen Güte des Fragebogens ermöglicht, sondern auch zu spezifischen Kausalzusammenhängen zwischen den Komponenten des CUE-Modells.

Um den vorliegenden Fragebogen auch als reine Papier-Bleistift-Variante einsetzen zu können, wird in einer aktuellen Studie die Übereinstimmung der Angaben im Modul 4 zwischen den verschiedenen Eingabeformen überprüft.

Fazit

Auf Basis eines integrativen Modells zur User Experience (CUE-Modell) wurde der *meCUE* Fragebogen zur modularen Evaluation zentraler Aspekte des Nutzungserlebens erfolgreich konstruiert. Er wurde theoriebasiert entwickelt und ist sowohl bei seiner Konstruktion als auch in zusätzlichen Studien unter Einsatz einer breiten Vielfalt unterschiedlicher interaktiver Produkte empirisch validiert worden (Minge & Riedel, accepted; Doria, Mingé & Riedel, accepted; Mingé, Riedel & Thüring, accepted). Damit bietet er eine ökonomische, valide und zuverlässige Alternative zu bereits bestehenden Verfahren. Insbesondere die modulare Struktur unter Berücksichtigung von Produktwahrnehmungen, Nutzeremotionen und Konsequenzen erlaubt es, quantitative Vergleiche zwischen verschiedenen Gestaltungslösungen oder zwischen verschiedenen Nutzungszeitpunkten zielorientiert und ganzheitlicher aus nutzerzentrierter Sicht vorzunehmen.

Literatur

- Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.
- DIN EN ISO 9241-210 (2011). Ergonomie der Mensch-System-Interaktion – Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme.
- Doria, L., Mingé, M. & Riedel, L. (accepted). User-centred evaluation of lower-limb orthoses: A new approach. Dreiländertagung der Deutschen, Schweizerischen und Österreichischen Gesellschaft für Biomedizinische Technik (BMT). 19.-21.09.2013, Graz.

- Hassenzahl, M. (2001). The effect of perceived hedonic quality on product appealingness. *International Journal of Human-Computer Interaction*, 13(4), 481-499.
- Hassenzahl, M., Burmester, M. & Koller, F. (2008). Der User Experience (UX) auf der Spur: Zum Einsatz von www.attrak.diff. In: H. Brau, S. Diefenbach, M. Hassenzahl, F. Koller, M. Peissner & K. Rose. (Hrsg.): *Usability Professionals 2008*, 78-82.
- Laugwitz, B., Schrepp, M. & Held, T. (2006). Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In A. M. Heinecke & H. Paul (Hrsg.), *M&C 2006: Mensch und Computer im Strukturwandel* (S. 125-134). München: Oldenbourg.
- Lavie, T. & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies*, 60, 269-298.
- Minge, M. & Riedel, L. (accepted). meCUE – Ein modularer Fragebogen zur Erfassung des Nutzungserlebens. Mensch und Computer, 08.-11.09.2013, Bremen.
- Minge, M., Riedel, L. & Thüning, M. (accepted). Und ob du wirklich richtig stehst... Zur diskriminativen Validität des User Experience Fragebogens „meCUE“. Beitrag im Workshop „Temporale Aspekte des Nutzungserlebens“, Mensch und Computer, 08.09.2013, Bremen.
- Thüning, M. & Mahlke, S. (2007). Usability, aesthetics, and emotions in human-technology interaction. *International Journal of Psychology*, 42(4), 253-264.
- Velicer, W.F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321-327.
- Watson, D., Clark, A. & Tellegen, A. (1988). Development and Validation of Brief Measure of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 54(6), 1063-1070.

Zur Validität eines User Experience-Fragebogens zur Messung der Bedürfniserfüllung in Produktinteraktionen

Moritz Körber¹ und Klaus Bengler¹

Schlüsselwörter: User Experience, Fragebogen, psychologisches Bedürfnis

Zusammenfassung

Psychologische Bedürfniserfüllung wurde bereits in mehreren Studien mit einer positiven Erfahrung bei Interaktionen mit Produkten in Verbindung gebracht. (z. B. Hassenzahl, Diefenbach & Göritz, 2010). Diese Studien beschränken sich meist allerdings nur auf berichtete Erlebnisse. Innerhalb von zwei Studien soll dieser Ansatz bei realen Produktinteraktionen untersucht werden. Dazu wurde ein Fragebogen, der UXNQ, zusammen mit dem AttrakDiff zur Bewertung neuartiger Interaktionskonzepte im Automobil benutzt. In beiden Studien stand die Bedürfniserfüllung mit dem positiven Affekt in einem positiven Zusammenhang, wobei es in Studie 2 auch eine Ausnahme gab. Sowohl die diskriminante als auch die konvergente Validität konnte anhand der Zusammenhänge mit dem AttrakDiff bestätigt werden. Die Studie zeigt, dass die Erfüllung psychologischer Bedürfnisse in Produktinteraktionen durch Design gezielt adressiert werden kann. Produktdesign, welches auf Bedürfnissen beruht, kann mit Hilfe einer Story positive Erlebnisse in einer Produktinteraktion schaffen. Die Wahl der Evaluationsmethode sollte allerdings auf den Entwicklungsstand abgestimmt sein.

Im Zuge des User Centered Designs wurde zur Produktevaluation bei den bisher starren Usability-Tests der Nutzungskontext immer stärker einbezogen und wurde nun zudem auf Freizeitaktivitäten erweitert (Boehm-Davis, Marcus, Green, Hada & Wheatly, 2003). Dies hat dazu geführt, dass Usability-Kriterien nicht mehr ausreichen, um die nicht-arbeitsrelevanten Aspekte von Produkten zu evaluieren. Neben der Usability wird deshalb die gesamte Erfahrung eines Nutzers mit einem Produkt, die User Experience, erfasst. Für den Begriff der User Experience existiert zwar Konsens hinsichtlich bestimmter Aspekte (Roto, Rantavuo & Väänänen-Vainio-Mattila, 2009), es existiert aber weder eine einheitliche Definition noch eine einheitliche Evaluationsmethode. Evaluationskriterien für User Experience beziehen sich beispielsweise auf Attribute des Produkts (z. B. visuelle Ästhetik; Lavie & Tractinsky, 2006), Produktfunktionen (Hassenzahl, Burmester & Koller, 2003) oder die subjektive Komponente von Emotionen (Bradley & Lang, 1994). Die Vielfalt von Methoden erfordert einen Nachweis, wie sich die einzelnen Methoden und damit gemessenen Konstrukte voneinander abgrenzen oder gegebenenfalls überschneiden. Eine Studie zur Validierung einer Methode muss daher einer-

¹ Technische Universität München, Lehrstuhl für Ergonomie, Boltzmannstraße 15, D - 85747 Garching, koerber@ife.mw.tum.de

seits zeigen, dass das vorgegebene Konstrukt gemessen wird, z. B. anhand eines empirischen Kriteriums. Andererseits muss ebenso gezeigt werden, dass die Methode nicht misst, was sie nicht messen soll. Fehlende Konstruktvalidität verhindert nicht nur die Entwicklung von User Experience-Modellen, sondern mindert auch den Wert eines Evaluationsergebnisses: Wenn nicht festgelegt ist, was eine Methode misst, dann können dementsprechend Designer nicht darlegen, ob die Designziele erfüllt wurden.

Theoretischer Hintergrund

Mehrere Autoren (Hassenzahl et al., 2010; Schulze & Krömker, 2010; Sproll, Preissner & Sturm, 2010) haben bereits hervorgehoben, dass es für ein positives Nutzungserlebnis wichtig ist, auf die Bedürfnisse der Nutzer zu achten. Der Bezug psychologischer Bedürfniserfüllung zu Wohlbefinden (Ryan & Deci, 2000), befriedigenden Erfahrungen (Sheldon, Elliot, Kim & Kasser, 2001) und positiven Nutzerberichten (Hassenzahl, 2008; Hassenzahl et al., 2010) wurde bereits nachgewiesen. Obwohl sich wie bereits erwähnt einige Studien mit der Bedürfniserfüllung befassen, existiert bisher wenig Forschung, bei der dieser Ansatz anhand einer realen Produktinteraktion untersucht wird. Deshalb wird in dieser Studie die Anwendung dieses Ansatzes in einer realen Produktinteraktion in zwei Studien untersucht. Der in dieser Studie verwendete Fragebogen *User Experience Need Questionnaire* (UXNQ) befindet sich noch in der Entwicklung und baut auf diesem theoretischen Ansatz auf: Mit dem Fragebogen lässt sich die Erfüllung von 9 psychologischen Bedürfnissen in einer Nutzungssituation messen. Der Fragebogen wurde anhand von drei Onlinestudien entwickelt und umfasst pro Bedürfnisskala 3–4 Items (Körper, Eichinger, Bengler & Olaverri, 2013), welche die Zustimmung zu bedürfnisrelevanten Aussagen auf einer 5-stufigen Likert-Skala erfassen. Der Nutzungskontext wurde durch einen *story telling* Ansatz (Michailidou, von Saucken & Lindemann, 2013) definiert. Der in Studie 1 verwendete Prototyp ist ein neuartiges Navigationsgerät, welches zwei Bedürfnisse ansprechen soll: Man kann damit einerseits auf der Fahrt Ziele entdecken, welche bei einer normalen Routenführung unentdeckt geblieben wären. Damit spricht das Gerät das Bedürfnis *Stimulation* an, welches für die faszinierte Entdeckung von Neuem und einem tollen Erlebnis, d.h. *Stimulation* durch das Leben im Gegensatz zu Langeweile und Routine, steht. Das Navigationsgerät kann andererseits außerdem an Mitfahrer weitergegeben werden, sodass sich dieser das ausgewählte Ziel selbst anschauen und gegebenenfalls ein neues Ziel aussuchen kann. Damit wird das Bedürfnis für *Verbundenheit* angesprochen, welches sich in einem Teamgefühl, einem Zugehörigkeitsgefühl oder einem Kontakt mit nahestehenden Menschen zeigt. Auf Grund dieses Designs wird erwartet, dass die Nutzung des Prototyps die Bedürfnisse "Verbundenheit" und "Stimulation" erfüllt. Da der Fokus von User Experience auf positiven Erfahrungen und damit nicht der Vermeidung von Frustration liegt (Hassenzahl & Tractinsky, 2006) und auf Grund bisheriger Forschung basierend

auf Nutzerberichten (Hassenzahl, 2008; Hassenzahl et al., 2010), wird erwartet, dass die Bedürfniserfüllung mit dem erlebten positiven Affekt zusammenhängt. In Studie 2 wird ein Prototyp verwendet, welcher den Start-Stop-Knopf eines Autos um eine Statusanzeige erweitert. Der Prototyp befand sich, im Gegensatz zum Prototyp aus Studie 1, allerdings noch in einer frühen Konzeptphase und besaß noch nicht die volle Funktionalität. Zur Überprüfung der Konstruktvalidität wurde mit dem Fragebogen AttrakDiff (Hassenzahl et al., 2003) ein weiterer User Experience Fragebogen herangezogen. Der AttrakDiff besteht aus den Dimensionen pragmatische Qualität, hedonische Qualität (Subskalen: Identität und Stimulation) und Attraktivität. Es wurde erwartet, dass die Bedürfniserfüllung unabhängig von der pragmatischen Qualität ist (diskriminante Validität). Ebenso sollte zur Subdimension Hedonische Qualität-Identität keine Korrelation bestehen Mit der Dimension Hedonische Qualität-Stimulation sollte die Skala Stimulation positiv korrelieren, da beide Skalen einen sehr ähnlichen Messgegenstand besitzen (konvergente Validität). Zur Attraktivität sollte, da sich diese aus der hedonischen und pragmatischen Qualität zusammensetzt, auch ein positiver, aber schwächerer Zusammenhang bestehen.

Methode

An Studie 1 nahmen insgesamt $n = 28$ Probanden teil, davon waren 19 (67.9 %) männlich und 9 (32.1 % weiblich). Das Alter lag im Mittel bei $M = 25.04$ ($SD = 6.09$). 4 Teilnehmer waren berufstätig, die restlichen 24 Teilnehmer waren Studierende. An Studie 2 nahmen insgesamt 29 Probanden teil, 25 (86.2 %) männlich, 4 (13.8 % weiblich), mit einem Alter im Mittel von $M = 25.18$ ($SD = 2.79$). Davon waren 5 Probanden berufstätig und 24 Probanden Studierende. Beide Studien weisen denselben Versuchsablauf auf: Nach einer Begrüßung wurde den Versuchspersonen per Storyboard die Story und damit der Nutzungskontext präsentiert. Die Versuchspersonen wurden angewiesen, sich in die handelnden Charaktere der Story hineinzuversetzen und sich vorzustellen, wie sie sich in der Situation fühlen würden. Die Story für Studie 1 handelte von einer Gruppe auf einer Freizeitfahrt auf der Autobahn. Sie kennen sich in der Gegend nicht aus und wollen mit Hilfe des Prototyps gemeinsam ein interessantes Ziel in der Umgebung entdecken. Da das Navigationsgerät eine gemeinsame Interaktion erfordert, nahmen immer zwei Personen gleichzeitig am Experiment teil. Anschließend sollten sich die Versuchspersonen in einen Demonstrator setzen und das Navigationsgerät wie in der Story als Bei- und Mitfahrer benutzen. Zur Messung der Bedürfnisse wurde der User Experience Need Questionnaire (UXNQ; Körber et al., 2013) benutzt. Aus den Bedürfnisskalen wurden die Skalen Stimulation und Verbundenheit ausgewählt. Für Studie 2 wurden die Bedürfnisse Kompetenz, Sicherheit und Stimulation ausgewählt. Als weiterer Hinweis auf den Zusammenhang zwischen Bedürfniserfüllung und eines positiven Erlebnisses wurde ein Interview durchgeführt, in welchem die Probanden gefragt wurden, was Ihnen an der Interaktion gefallen hat. Zur Messung des positiven Affekts wurde die Kurzversion des PANAS (Kercher, 1992), über-

setzt von Krohne, Egloff, Kohlmann und Tausch (1996), verwendet. Der PANAS-Fragebogen besteht aus den Subskalen „Positiver Affekt“ (PA; angenehme Erregung, Euphorie) und „Negativer Affekt“ (NA; unangenehme Erregung, Stress). Der AttrakDiff besteht aus den Skalen „Pragmatische Qualität“ (PQ; im Wesentlichen die Gebrauchstauglichkeit), „hedonische Qualität-Identität“ (HQ-I; inwieweit dient das Produkt dazu, anderen meine Identität zu kommunizieren), hedonische Qualität-Stimulation (HQ-S; Anregung durch neue Interaktionsstile oder Funktionalitäten) und Attraktivität (ATT; globales Produkturteil).

Ergebnisse

In Tabelle 1 sind die deskriptiven Kennwerte der zwei Bedürfnisskalen-Skalen gelistet. Ein Bedürfnis wurde als erfüllt bewertet, wenn sein Mittelwert signifikant über der Skalenmitte von 3 lag. Dies wurde mit Hilfe eines t-tests für eine Stichprobe überprüft. Sowohl Verbundenheit ($t(27) = 3.65; p < .001$) als auch für Stimulation ($t(26) = 9.880; p < .001$) können in der Interaktion als erfüllt bewertet werden. Um den Zusammenhang zwischen der Bedürfniserfüllung und dem positiven Erleben zu untersuchen, wurde die bivariate Korrelation nach Pearson ermittelt. Die Ergebnisse sind in Tabelle 2 gelistet. Wie erwartet hingen die Bedürfnisse Verbundenheit ($r = .41, p < .05$) und Stimulation ($r = .45, p < .05$) signifikant mit dem erlebten positiven Affekt zusammen. Es bestand kein Zusammenhang zum negativen Affekt ($r = -.15, p = n. s.$ bzw. $r = -.21, p = n. s.$). In Studie 2 sollte die diskriminante und konvergente Validität überprüft werden.

Tab.1: Deskriptive der zwei Bedürfnisskalen und der PANAS-Skalen.

	M	SD	α	t	df	Effektstärke r
Verbundenheit	3.56	.81	.69	3.648	27	.57
Stimulation	4.07	.56	.62	9.880	26	.89
PA	3.37	.72	.77			
NA	1.27	.27	.48			

Anmerkungen. PA = positiver Affekt, NA = negativer Affekt.

Tab.2: Zusammenhang zwischen der Bedürfniserfüllung und dem PANAS

	Verbundenheit	Stimulation	PA	NA
Verbundenheit	(.69)			
Stimulation	.36	(.62)		
PA	.41*	.45*	(.77)	
NA	-.17	-.21	.20	(.48)

Dazu wurden Korrelationen zu einem anderen User Experience Fragebogen, dem AttrakDiff, durchgeführt. Die Ergebnisse sind in Tabelle 3 aufgeführt. Wie erwartet bestand kein Zusammenhang zwischen der wahrgenommenen pragmatischen Qualität des Prototypen und den Bedürfnissen Kompetenz ($r = .27, p = n. s.$), Sicherheit ($r = .13, p = n. s.$) oder Stimulation ($r = .15, p = n. s.$). Ebenfalls gemäß den Erwartungen war die signifikante Korrelation zwischen Stimulation und HQ-S ($r = .72, p < .01$) und zwischen Stimulation und der Attraktivität ($r = .61, p = .01$). Der Unterschied zwischen diesen beiden Korrelationen war nicht signifikant ($z = .49, p = n. s.$). Dieser Zusammenhang bestand nicht für die anderen beiden Bedürfnisse Kompetenz ($r = .04, p = n. s.$ bzw. $r = .30, p = n. s.$) und Sicherheit ($r = .13, p = n. s.$ bzw. $r = .03, p = n. s.$). Entgegen den Erwartungen korrelierten die als unabhängig deklarierten Skalen HQI und PQ ($r = .61, p < .05$) sowie HQ-S und HQ-I ($r = .75, p < .01$) signifikant. Ebenfalls entgegen der deklarierten Struktur des Fragebogens ist die Korrelation der Skalen Sicherheit und Kompetenz ($r = .77, p < .01$) und Stimulation mit HQ-I ($r = .57, p < .05$). Die Bedürfnisse Kompetenz ($r = .56, p < .01$) und Stimulation ($r = .55, p < .01$) standen in einem positivem Zusammenhang mit der positiven Affekt-Skala. Dies galt nicht für das Bedürfnis Sicherheit ($r = .26, p = n. s.$). Zur negativen Affekt-Skala bestand kein signifikanter Zusammenhang ($r = -.25, p = n. s.$ bzw. $r = -.06, p = n. s.$ bzw. $r = -.01, p = n. s.$). Die Ergebnisse sind in Tabelle 4 gelistet.

Tab.3: Zusammenhang zwischen dem UXNQ und dem AttrakDiff

	K	S	Stim	PQ	HQS	HQI	Att
K	(.72)						
S	.77**	(-.34)					
Stim	.26	-.26m	(.63)				
PQ	.27	.13	.15	(.71)			
HQS	.04	-.048	.72**	.31	(.58)		
HQI	.29	-.293	.57*	.61*	.75**	(.84)	
ATT	.30	.03	.61*	.74**	.77**	.87**	(.90)

Anmerkungen. K = Körperlichkeit, S= Sicherheit, Stim = Stimulation.

Tab.4: Zusammenhang zwischen UXNQ und PANAS

	K	S	Stim	PA	NA
Kompetenz	(.72)				
Sicherheit	.77**	(-.34)			
Stimulation	.26	-.23	(.63)		
PA	.56**	.26	.55**	(.76)	
NA	-.25	-.06	.01	.27	(.34)

Diskussion

Ziel dieser zwei Studien war es, die Validität des UXNQ anhand einer realen Interaktion mit einem Prototyp zu überprüfen. In Studie 1 wurde ein Navigationsgerät evaluiert, welches die Bedürfnisse Verbundenheit und Stimulation in der Interaktion erfüllen sollte. Wie erwartet gaben die Probanden im Fragebogen die Bedürfnisse als in der Interaktion erfüllt an. Dies zeigte sich auch in den Kommentaren der Nutzer: 17 der 28 Teilnehmer nannten auf die Frage "Was hat Ihnen gefallen?" entweder die Interaktion mit den anderen Mitfahrern als Team (Verbundenheit; z. B. "Das soziale Erlebnis: Spielerisch wird die Umgebung gemeinsam entdeckt") oder den Spaß, etwas Neues zu entdecken (Stimulation; "Tolle neue Erfahrung während des Autofahrens (keine Langeweile)"). Die Interaktion mit den Prototypen wurde im PANAS als positives Erlebnis gewertet. Wie erwartet ergaben sich ein positiver Zusammenhang zwischen der Bedürfniserfüllung und der positiven Affektskala des PANAS. Dies sind erste Hinweise, dass eine Bedürfniserfüllung auch in realen Interaktionen mit einer positiven Erfahrung in Verbindung steht. Dieser Befund konnte allerdings in Studie 2 nur teilweise repliziert werden, da für eine Bedürfnisskala (Sicherheit) keine signifikante Korrelation gefunden wurde. In Studie 2 wurde der neben dem UXNQ zur Evaluation auch der AttrakDiff benutzt, um Überschneidungen und Abgrenzungen mit diesem Fragebogen zu erfassen. Wie erwartet bestand ein hoher Zusammenhang der Bedürfnisskala Stimulation mit der Skala hedonische Qualität-Stimulation sowie mit der Attraktivität. Entgegen den Erwartungen waren allerdings die zum Teil sehr hohen Interkorrelationen zwischen den Skalen innerhalb beider Fragebogen und die mitunter geringe interne Konsistenz der Skalen. Beide Befunde könnten durch die Nutzungssituation erklärt werden: In Studie 2 wurde ein Prototyp im sehr frühen Entwicklungsstadium getestet. Zudem wurde kein Fahrsimulator genutzt, um eine reale Fahrsituation zu simulieren, weshalb sich die Probanden einen großen Teil aus der in der Story beschriebenen Fahrsituation vorstellen mussten. Dies ist allerdings im Sinne der agilen Entwicklung nicht ungewöhnlich, da hier schnell erste Ergebnisse geliefert werden müssen, um diese iterativ in die Entwicklung einfließen zu lassen. Der UXNQ beinhaltet sehr abstrakte Aussagen, welche teilweise auch invers formuliert sind. Somit sind die Items möglicherweise auf Grund des unausgearbeiteten Nutzungskontextes für die Versuchspersonen schwer zu beantworten oder gar verwirrend. Der UXNQ scheint seinen Einsatzzweck also primär in einer detaillierten Studie zu haben, welche einen realitätsnahen Nutzungskontext bietet, der möglichst nah an der Produkt-Story liegt. Dies zeigt gleichzeitig, wie wichtig der Einsatz einer Story und damit eines Nutzungskontext ist, um ein Erlebnis zu transportieren oder zu generieren. Für Evaluationen im frühen Entwicklungsstadium, in dem der Nutzungskontext nur sehr vage dargestellt werden kann, könnten deshalb Expertenevaluationen oder simplere Messmethoden sowie Interviews nützlichere Informationen bieten. In dieser Studie wurden nur korrelative Zusammenhänge untersucht. Experimentelle Designs können die Rolle von Bedürfnissen in Interaktionen weiter

aufschlüsseln. Für eine summative Evaluation sollten außerdem Teilnehmer aus der Zielgruppe anstelle von Studenten als Versuchsteilnehmer eingeladen werden.

Literatur

- Boehm-Davis, D. A., Marcus, A., Green, P. A., Hada, H., & Wheatley, D. (2003). The Next Revolution: Vehicle User-Interfaces and the Global Rider/Driver Experience. In *Extended abstracts of the 2003 Conference on Human Factors in Computing Systems, CHI '03* (pp. 708–709).
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *Journal of behavior therapy and experimental psychiatry*, 25(1), 49–59.
- Hassenzahl, M. (2008). User experience (UX): towards an experiential perspective on product quality. In *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine* (pp. 11–15). New York, NY: ACM.
- Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In G. Szwillus & J. Ziegler (Eds.), *Mensch & Computer 2003* (pp. 187–196). Wiesbaden: Vieweg+Teubner Verlag.
- Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products – Facets of user experience. *Interacting with Computers*, 22(5), 353–362. doi:10.1016/j.intcom.2010.04.002
- Hassenzahl, M., & Tractinsky, N. (2006). User experience – a research agenda. *Behaviour & Information Technology*, 25(2), 91–97. doi:10.1080/01449290500330331
- Kercher, K. (1992). Assessing Subjective Well-Being in the Old-Old: The PANAS as a Measure of Orthogonal Dimensions of Positive and Negative Affect. *Research on Aging*, 14(2), 131–168. doi:10.1177/0164027592142001
- Körber, M., Eichinger, A., Bengler, K., & Olaverri-Monreal, C. (2013). User Experience Evaluation in an Automotive Context. In *Proceedings of the 2013 International IEEE Intelligent Vehicles Symposium Workshops*.
- Krohne, H. W., Egloff, B., Kohlmann, C.-W., & Tausch, A. (1996). Untersuchung mit einer deutschen Form der Positive and Negative Affect Schedule (PANAS). *Diagnostica*, 42, 139–156.
- Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies*, 60(3), 269–298. doi:10.1016/j.ijhcs.2003.09.002

Engineering meet UX: Ingenieurstechnische Absicherungsmethoden treffen auf User Experience Forschung

Maik Auricht¹, Elisabeth Dittrich², Stefan Brandenburg³ und Rainer Stark⁴

Schlüsselwörter: User Experience, Systems Engineering, Validierung, Produktentstehungsprozess, Systemsimulation, interaktive Prototypen

Zusammenfassung

Um den Praxisbedarf der erlebbaren Absicherung von Prototypen in frühen Entwicklungsphasen zu verdeutlichen und Ansätze für die Integration von Nutzertests in moderne Vorgehensmodelle zu evaluieren, wurde eine Fragebogenuntersuchung mit 36 Experten durchgeführt. Diese Experten stammten aus dem Bereich der Virtuellen Produktentstehung und wiesen einen ingenieurstechnischen Hintergrund auf. Der vorliegende Beitrag stellt die Ergebnisse der Befragung vor. Die dabei gewonnenen Erkenntnisse werden vor dem Hintergrund der Integration von Usability und User Experience Maßnahmen in den Produktentwicklungsprozess diskutiert und sollen helfen, die Absicherungen technischer Systeme zu verbessern.

Motivation und Hintergrund

Die Entwicklung technischer Systeme hat sich im Laufe der Jahre stark geändert: Wurden technische Systeme früher maßgeblich durch Ausprobieren und sukzessives Zusammenbauen hergestellt, so ist der Produktentstehungsprozess im Laufe der Zeit digitalisiert worden. Durch die rapide Veränderung der Informationstechnik hat sich neben der Produktentwicklung auch deren Absicherung in den digitalen Bereich verschoben. Die Absicherung mechatronischer Systeme, d. h. deren frühzeitiger Test, spielt dabei eine zentrale Rolle im gesamten Produktentstehungsprozess (Albers, 2010). Innerhalb dieser Systemabsicherung existieren wiederum diverse Methoden, um digitale Modelle auf ihre Funktionsweise zu überprüfen (Model/Software/Hardware-in-the-Loop, Multiphysiksimulationen, Finite-Elemente-Methoden, hybride und reale Prototypen usw.). Dabei kommen Prototypen mit physischen Elementen erst sehr spät im Entwicklungsprozess zum Einsatz. Entsprechend sind direkte ergonomische oder funktionale Untersuchungen, die eine Erlebbarkeit des Prototypen möglich machen, im frühen Entwicklungsprozess bisher kaum berücksichtigt (vgl. Stork et al., 2010). Dabei erfordert nicht nur die Funktionalität eines technischen Systems sondern auch die Qualität der Interaktion von Mensch und Maschine eine nutzerzentrierte Absicherung. Bei einem derarti-

¹ Technische Universität Berlin, Sekr. PTZ 4, Pascalstraße 8-9, 10587 Berlin, maik.auricht@tu-berlin.de

² Graduiertenkolleg Prometei, Sekr. MAR 3-3, Marchstraße 23, 10587 Berlin, elisabeth.dittrich@zmms.tu-berlin.de

³ Technische Universität Berlin, Sekr. MAR 3-2, Marchstraße 23, 10587 Berlin, stefan.brandenburg@tu-berlin.de

⁴ Fraunhofer IPK, Sekr. PTZ 4, Pascalstraße 8-9, 10587 Berlin, rainer.stark@ipk.fraunhofer.de

gen Test sollten zwei Aspekte bedacht werden: Die Gebrauchstauglichkeit des Produkts (Usability) und das Erleben der Nutzer (User Experience). Während sich die Usability eines Produktes auf seine instrumentellen Qualitäten (bspw. Grad der Zielerreichung) bezieht, fokussiert die User Experience dessen nicht-instrumentellen Eigenschaften, wie z. B. Ästhetik oder Form (Mahlke & Thüring, 2007). Beide Aspekte sind nicht unabhängig von einander (vgl. Hassenzahl, 2003) und beeinflussen gemeinsam die Akzeptanz des technischen Artefakts (Mahlke & Thüring, 2007; Norman, 2004). Daher sollten technische Systeme in frühen Phasen der Produktentwicklung auch auf deren Gebrauchstauglichkeit und Wirkung auf das emotionale Erleben der Nutzer getestet werden. In wie weit dies bereits geschieht und wo Verbesserungsbedarf besteht, arbeitet die vorliegende Untersuchung mit einer Expertenumfrage heraus. Dabei werden die folgenden Forschungsfragen beantwortet:

- In welchen Bereichen der Produktentwicklung werden digitale, interaktive Prototypen eingesetzt?
- Inwieweit werden Methoden und Verfahren aus dem Usability Engineering (UE) und der User Experience (UX) Forschung für die Absicherung von interaktiven Prototypen verwendet?
- Wenn es Verbesserungsbedarf hinsichtlich der Absicherung technischer Systeme gibt, wie sollte eine derartige Absicherung idealerweise aussehen?

Methode

Stichprobe: An der Untersuchung nahmen 36 Personen teil. Im Mittel waren diese 30.40 Jahre alt ($SD=4.41$). Alle waren in Forschung und Entwicklung im Bereich der Produktentstehung tätig. 61% gaben an, mittelmäßig oder viel Erfahrung mit Projekten im Bereich mechatronischer Systeme zu haben. Weiterhin verfügten jeweils 40% der Teilnehmer über mittel bis viel Erfahrung mit der Simulation von Systemen und in der Arbeit mit Prototypen.

Material: Um die gestellten Forschungsfragen zu beantworten, wurde ein Online-Fragebogen mit insgesamt 18 Fragen in Soscisurvey (www.soscisurvey.de, letzter Abruf 15.07.2013) entworfen. Die Items waren jeweils aus einer von drei Kategorien: 1) 7 für die Erfassung der Erfahrung mit Absicherungsprozessen, 2) 7 aus dem Bereich User Experience und 3) weitere vier Fragen zu demographischen Aspekten. 10 der 18 Fragen hatten ein freies Antwortformat. Für weitere 2 Fragen konnten die Versuchspersonen entweder ein 4-stufiges semantisches Differenzial mit den Verankerungen „stimme gar nicht zu“, „stimme teilweise nicht zu“, „stimme teilweise zu“ und „stimme voll zu“ zur Beantwortung nutzen. Für die verbleibenden 6 Fragen wählten die Probanden aus fest vorgegebenen Antworten aus. Das Ausfüllen des Fragebogens dauerte ca. 15 Minuten.

Ergebnisse

Aufgrund der Fülle der Fragebogenergebnisse können nachfolgend nur ausgewählte Ergebnisse zur Beantwortung der Forschungsfragen aufgeführt werden. Nicht gegebene Antworten innerhalb der Fragen werden nicht mit aufgeführt.

Zur Beantwortung der 1. Forschungsfrage, in welchen Bereichen interaktive Prototypen eingesetzt werden, ergibt sich folgendes Bild. Domänenübergreifend setzen 26% der Befragten interaktive Prototypen ein. Weiterhin antworteten 26%, dass sie diese Art von Prototypen bereits in frühen Phasen der Produktentwicklung nutzen. Jeweils 11% der Befragten antworteten, dass sie interaktive Prototypen in der Softwareentwicklung, der Fabrikplanung und der Ergonomieabsicherung verwenden. 37% der Teilnehmer setzten gar keine Prototypen ein. Die Frage "In welchen Bereichen sind interaktive Prototypen für Sie sinnvoll?" beantworteten 31% der Befragten mit in der Absicherung allgemein. 25% sagten innerhalb der Entwicklung, gefolgt von 22% bei Interfaces/Bedienelementen. 19% gaben an, dass sie interaktive Prototypen im Bereich der Softwareentwicklung sinnvoll finden. Weitere 16% bewerteten Prototypen in der Mechatronikentwicklung als bedeutsam. Demnach setzt eine Mehrzahl der Befragten interaktive Prototypen in der Technikentwicklung ein und findet diesen Einsatz auch sinnvoll. Unterstützt wird diese Schlussfolgerung unter anderem dadurch, dass viele der Befragten die Nichterfüllung von UE und UX Bestandteilen für Akzeptanzprobleme beim Anwender mechatronischer Systeme verantwortlich machten (siehe Tab. 1) und eigene schlechte Erfahrungen mit mechatronischen Systemen hauptsächlich auf schlechte Usability (76%) und negative Emotionen bei der Interaktion (37%) zurückführten. Technische Probleme waren nur für 8% der Befragten Ursache für Akzeptanzprobleme.

Tabelle 1: Antworten zur Frage "Warum kommt es Ihrer Meinung nach zu Akzeptanzproblemen beim Anwender mechatronischer Systeme?"

Kategorie	Erwartungskonformität nicht erfüllt	Selbstbeschreibungsfähigkeit nicht erfüllt	Steuerbarkeit nicht erfüllt	Systemvertrauen nicht erfüllt	fehlende erlebbare Absicherung
Anzahl Antworten	12	13	15	14	8
Verteilung in %	15%	17%	19%	18%	10%

Hinsichtlich der zweiten Forschungsfrage, inwieweit Methoden und Verfahren aus dem UE und der UX Forschung für die Absicherung von interaktiven Prototypen verwendet werden, antworteten 76% der befragten Personen, dass der Einsatz von interaktiven Prototypen in der Produktentwicklung zu spät zum Einsatz kommt (Abb. 1). 76% lehnten ebenfalls die Aussage ab, dass der Einsatz von interaktiven Prototypen keiner Verbesserung bedarf. 93% antworteten, dass die Emotionen für

die Akzeptanz eines Produktes bzw. einer Funktion wichtig sind. Die Funktionen mechatronischer Systeme nur am Computer abzusichern, verneinten 86%.

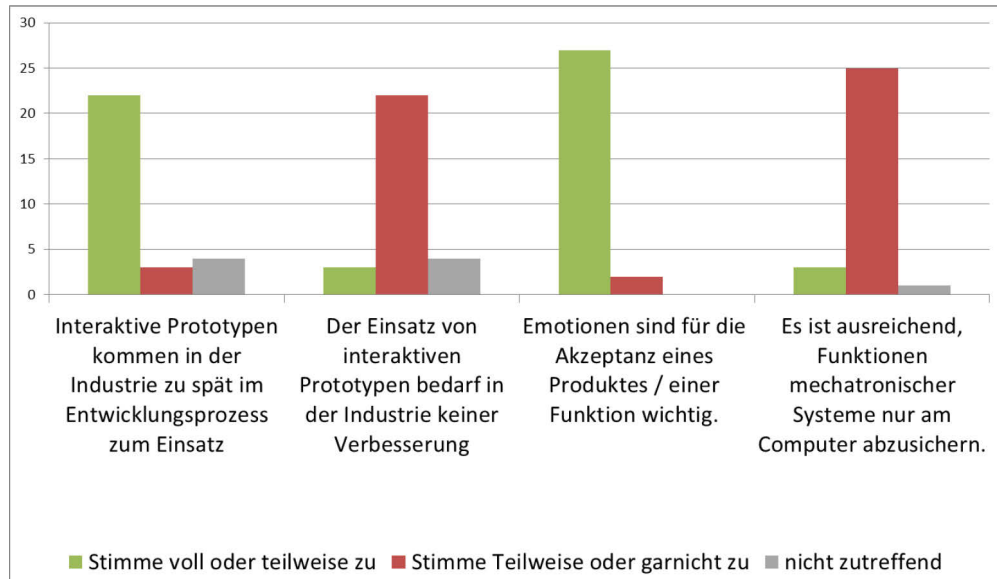


Abb. 1: Aufschlüsselung der Antworten zur Rolle interaktiver Prototypen bei der Produktentwicklung. Die Höhe der Balken entspricht der Häufigkeit der Antworten.

Auf die Frage, welche User Experience Methoden die Befragten bereits angewendet haben, gaben die meisten Teilnehmer an, noch keine UX-Methoden genutzt zu haben (Tab. 2).

Tabelle 2: Antworten zur Frage "Welche User Experience Methoden haben Sie bereits angewendet?"

Kategorie	Keine	3D Visualisierung	Haptische Prototypen	Customer Experience Interaction Model
Anzahl Antworten	13	2	1	1
Verteilung der Antworten in %	76%	12%	6%	6%

Dementsprechend eindeutig fielen auch die Antworten zur Frage aus, welche Methoden sich eignen, um mechatronische Systeme hinsichtlich deren Akzeptanz zu bewerten. Hier gaben 67% der Befragten keine Antwort an. Jeweils 11% hielten den AttrakDiff, den NASA-TLX und Augmented Reality für geeignet. Zusammenfassend wurde deutlich, dass UE und UX Ansätze in der Industrie zwar als wichtig erachtet, jedoch zu selten eingesetzt werden.

Bezüglich der dritten Forschungsfrage, nach dem (UE- & UX-) Verbesserungsbedarf bei der Absicherung digitaler Prototypen, eignen sich die Antworten zur Frage "Welche Wahrnehmungsbereiche sind Ihrer Meinung nach für die erlebbare Absicherung wichtig?". Dabei gaben 100% der Befragten an, dass die Wahrnehmung von Bewegung und die dreidimensionale visuelle Wahrnehmung wichtig

sind. 97% kreuzten an, dass die Wahrnehmung von Geräuschen und Bewegung wichtig ist. Sogar Gerüche und Geschmacksreize wurden von einem Teil der Befragten als wichtig erachtet (Abb. 2).

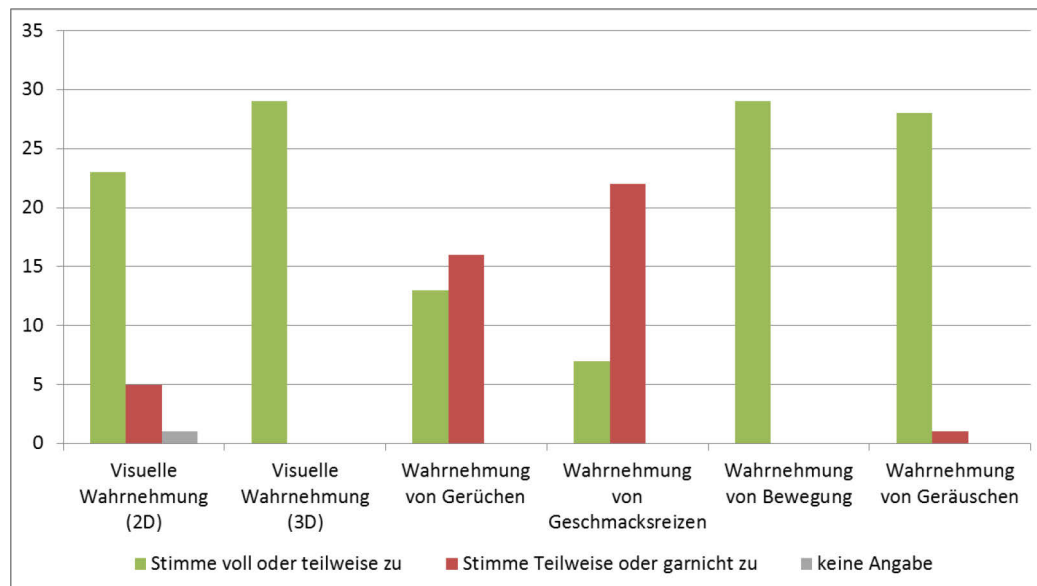


Abb. 2: Grafische Auswertung der Antworten zur Frage "Welche Wahrnehmungsbereiche sind Ihrer Meinung nach für die erlebbare Absicherung wichtig?" Die Höhe der Balken entspricht der Häufigkeit der Antworten.

Hinsichtlich des Verbesserungspotentials beim Einsatz von interaktiven Prototypen würden 95% der Befragten einen Wizard bevorzugen, der ihnen bei dem Transferieren des Simulationsmodells in die Versuchsumgebung hilft. 29% gaben hierbei an, dass der Aufbau dieses interaktiven Prototypen 1-2 Tage dauern kann. 14% würden auch 4-5 Tage tolerieren. Insgesamt existieren nach Ansicht der befragten Ingenieure also durchaus Verbesserungsmöglichkeiten hinsichtlich der Implementation von interaktiven Prototypen und deren erlebbarer Absicherung.

Diskussion

Die vorliegende Arbeit hatte das Ziel die Perspektive der Produktentwickler und die der UE/UX Forscher zusammenzubringen. Dafür wurde eine Umfrage erstellt und drei Forschungsfragen formuliert. Bezüglich der ersten Forschungsfrage wurde deutlich, dass die befragten Ingenieure interaktive Prototypen für wichtig halten und diese auch in der Produktentwicklung einsetzen. Darüber hinaus fanden sie die Einbindung von Maßnahmen des Usability Engineering und der User Experience Forschung in die Absicherung von interaktiven Prototypen wichtig. Deutlich wurde in den Antworten zur Forschungsfrage zwei allerdings auch, dass derartige Maßnahmen bisher kaum angewendet werden. Möglicherweise sind die Methoden und Verfahren des UE und UX bei Entwicklern nicht bekannt genug, um in der Produktentwicklung eingesetzt zu werden. Zumindest einige Antworten zum Einsatz von UE/UX Maßnahmen unterstützen diese Schlussfolgerung. Schlussendlich führte die dritte Forschungsfrage auf, dass Produktentwickler hauptsächlich auf

visuelle, auditive und kinästhetische Informationen bei der Evaluation von interaktiven Prototypen setzen. Damit stimmen die Ergebnisse der Umfrage mit Befunden aus verwandten Domänen überein (z. B. Kessner et al., Woywode et al., 2011). Zwar beschäftigten sich beide Untersuchungen mit der Integration von UE/UX Maßnahmen in den Softwareentwicklungsprozess bei kleinen und mittelständischen Unternehmen (KMU), doch fanden sie übereinstimmend, dass dort vergleichbare Probleme bei der Zusammenführung der Ingenieurs- und der UE/UX-Perspektiven entstehen. Zumeist wurden UE/UX Maßnahmen von den Produktentwicklern als wichtig erachtet, jedoch fehlte oft das Wissen über einzelne Methoden bzw. die Ressourcen sich dieses anzueignen. Um diese Probleme anzugehen, entwickelten Reckin und Brandenburg (im selben Band) Kriterien, die ein Modell des Usability Engineering erfüllen muss, um derartige Probleme bei der Einführung von UE/UX Maßnahmen in den Produktentstehungsprozess zu vermeiden (vgl. auch Stade et al., 2013).

Zukünftige Forschung sollte sich weiterhin mit der Einführung von UE/UX Maßnahmen in die erlebbare Absicherung von technischen Systemen beschäftigen. Nicht zuletzt zeigen die Ergebnisse dieser Untersuchung dahingehend einen großen Bedarf bei den Ingenieuren auf.

Literatur

- Albers, A. (2010). Five Hypotheses about Engineering Processes and their Consequences. In *Proceedings of the TMCE 2010*. Ancona.
- Hassenzahl, M. (2003). The Thing and I: Understanding the Relationship Between User and Product. In Mark A. Blythe, Andrew F. Monk, Kees Overbeeke und Peter C. Wright (Hrsg.). *Funology: From Usability to Enjoyment*, 1—12: Kluwer Academic Publishers.
- Kessner, D., Dittrich, F., & Bär, N. (2011). Von der Wissenschaft in die Wirtschaft - Wissenstransfer in Sachen Usability. In *Usability Professionals 2011*.
- Mahlke, S., & Thüring, M. (2007). Studying antecedents of emotional experiences in interactive contexts. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (S. 915-918). ACM.
- Norman, D. A. (2007). Emotional design: Why we love (or hate) everyday things. Basic books.
- Stade, M., Reckin, R., Brandenburg, S., & Thüring, M. (2013). Usability in KMU etablieren: Von schneller Problemlösung zu ressourcenorientiertem Usability Engineering. Paper auf der Mensch und Computer. 09-11 September, Bremen.
- Stork, A; Wagner, M., Schneider, P., Hinnerichs, A., Bruder, T., Eufinger, J. (2010). FunctionalDMU: Co-Simulation mechatronischer Systeme in einem DMU-Umfeld. In *Produktdaten-Journal (Nr. 1)*. S.44-48.

Woywode, M., Mädche, A., Wallach, D., & Plach, M. (2011). Gebrauchstauglichkeit von Anwendungssoftware als Wettbewerbsfaktor für kleine und mittlere Unternehmen (KMU) (Abschlussbericht). Institut für Mittelstandsforschung an der Universität Mannheim; Lehrstuhl für Wirtschaftsinformatik IV der Universität Mannheim; Fachhochschule Kaiserslautern; ERGOSIGN GmbH

Non-intrusive systems for human state diagnosis based on facial recognition and eye-tracking systems

Philippe Polet¹, M. Rötting², Catholijn Jonker³, Peter A. Wieringa⁴ and Frederic Vanderhaegen¹

Zusammenfassung

Dieses Papier schlägt eine Diskussion über die Bedeutung von nicht-intrusive Systeme vor, um den Zustand des menschlichen Bediener zu diagnostizieren. Diese Systeme lassen sich in verschiedenen Technologien zusammen: System-Analyse von Mimik, Augenbewegungen, fundierte Analyse. Mehrere Teams in der European Research Network HAMASYT beteiligt haben Kompetenzen in diesem Bereich entwickelt.

Diese Analyse des menschlichen Bediener Verfahren haben mehrere Vorteile: Der erste Monitor / Diagnose des Zustands des menschlichen Bedieners, erkennt der zweite abnormale Situationen.

Überwachung des Zustands des menschlichen Bedieners kann sich auf seine Wachsamkeit, sein Arbeitspensum, seinen emotionalen Zustand. Es ist auch möglich, das Verhalten der Bedienungsperson durch die Untersuchung seiner Aktionen auf dem System oder dessen Bewegung überwacht wird.

Introduction

The analysis and the control of Human-Machine Systems (HMS) depend on the observability of the different components of the system. As a component of the system the human operator state has to be observed in order to diagnose the system state. In this paper we are interesting in non-intrusive systems for human state diagnosis. Information used comes from non-intrusive sensors such as eye-tracking system or video of the human operator face. Two main reasons lead us to diagnose the human operator state. The first one is to verify if he/she behaves as required.

- In terms of activity: usually a set of activities are prescribed to the human operator. For instance the human operator is supposed to follow a specific procedure (list of action to be realized). Some sensors and tools can be used to analyze if the human operator respects the different steps of a procedure (Wegerich, et al. 2013). The eyes movements can be analyzed in order to de-

¹ Laboratory of Industrial and Human Automation, Mechanics and Computer Science, LAMIH, CNRS UMR 8201, University of Valenciennes, Le Mont Houy, F-59313 Valenciennes Cedex 9, France

² Berlin Institute of Technology, Department of Psychology and Ergonomics, Chair of Human-Machine Systems, Sekr. MAR 3-1, Marchstr. 23, 10587 Berlin

³ Delft University of Technology Faculty of Electrical Engineering, Mathematics and Computer, Mekelweg 2, 2628 CD Delft, The Netherlands

⁴ Delft University of Technology, Faculty of Mechanical Engineering, BioMechanical Engineering, Mekelweg 2, 2628 CD Delft, The Netherlands

termine what is watched by the human operator (Loslever, et al. 2008). It is then possible to verify if the human operator does the required checks.

- In terms of state of the human operator: some sensors and tools can be used in order to determine in which state is the human operator (Popieul, et al. 2012). Hypo-vigilance state can be determined by observing gaze eye movement and facial analysis (Venjakob, et al. 2012). Observation of the human operator can give some information about overload in terms of workload (some tasks neglected, stop of talking, etc...).

The second reason concerns the possibility to use the human operator as a sensor. When a problem is detected by a human operator, when he/she surprised by a particular event, some of his/her characteristics may change. For instance, they are characteristics changes related to facial expression, verbal communication (onomatopoeia), or specific gestures. The observation of the human operator can be used to detect some dangerous events, or difficulties to solve some problem. The observation of the human operator becomes then an indirect mean to diagnose the overall system (HMS + environment). Eye tracking and facial expression recognition can be used.

We consider only non-intrusive way to observe the human-operator in order to make easy the operationalization of our observation system. Different information can be used to observe the human operator:

- Human-Machine interface analysis (HMI): information from HMI is useful to know actions made by the human operator
- Image analysis: can be used to observe gesture, to analyze facial expression, etc. Such use of images from camera was used to detect human emotion and affect (Heloir et Kipp 2010) (Broekens, Qu et Brinkman 2012). Some of these approaches are applied for face animation of virtual agent in order to simulate emotions. Their competences are useful for this research topic by applying the inverse process (i.e., in order to recognize the emotional state of a human operator).
- Sound analysis: the sound level, speech recognition may give information about intention of the human operator or his/her emotional state.
- Eye tracking system analysis: to be used in order to know what is (or not) watched by the human operator. The characteristics of eyes, gaze can be used to determine hypo-vigilance state. Some researchers used such an approach to assess hypovigilance of human (Popieul, et al. 2012) or strategies of information processing (Venjakob, et al. 2012)), or human activity analysis (Onnasch, Ruff et Manzey 2012)

Application to railway driver analysis

A preliminary study has been evolved in order to highlights the characteristics of a railway driver. This study was realized with a railway simulator called COR&GEST (cf. Figure 1). COR&GEST is composed with a railway network (a) and the human-machine interface (HMI) for the driver (b).



Figure 1 (a) The railway network

(b) The human-machine interface.

This preliminary experiment focused on several kinds of data in order to classify the driver into 3 classes: hyper-reactive, normal, hypo-reactive.

This classification is based on:

- Behavior observed of the driver: the behavior is observed through the train movement. The main variable measurable is the train speed. Two main characteristics can then be studied:
- The time delay of the driver: the driver must respect a speed reference. The time delay is measured when a change of speed reference occurs.
- The precision: correspond to the gap between speed reference and train speed.
- Eye tracking measures:
- Eye clipping
- Eye movement; determine where the driver is looking.

Moreover, some alarms occurred during the driving. Reactions of drivers after these disturbances were analyzed. The goal of this paper is not to present detailed results of this preliminary study, but this section aims at illustrating some of observed results. Figures 2 and 3 illustrate the behavioral observation two drivers.

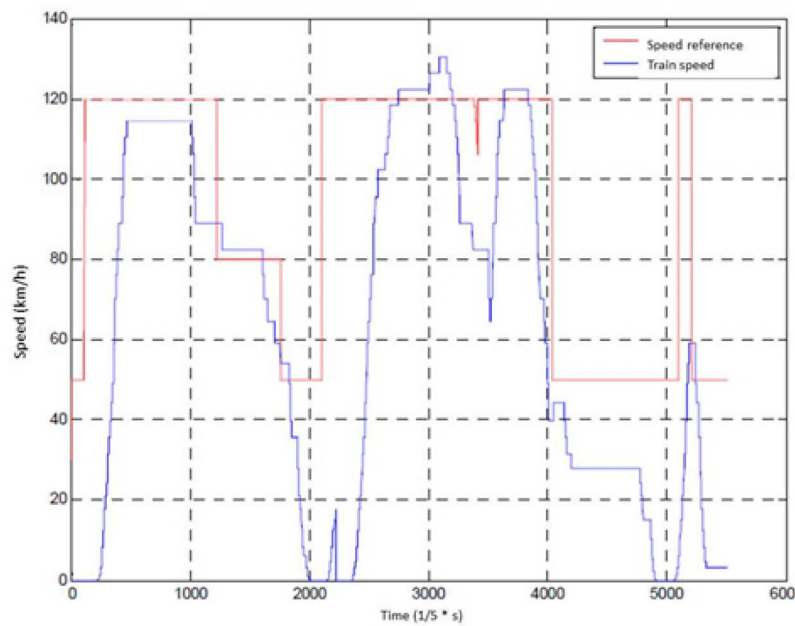


Figure 2 Behavior of the first driver.

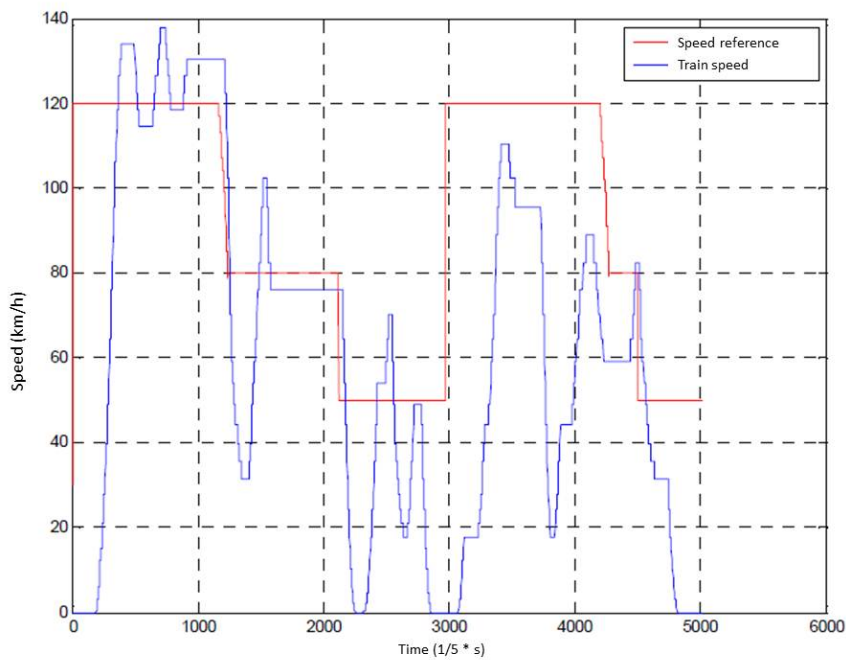


Figure 3 the behavior of the second driver.

The first driver on this sequence gives a “normal” behavior. The precision is not always correct. The subject miss understood very-low speed reference. In this case 3 squares were displayed signifying that the speed train must to be below 50 km/h (on sight driving). The driver interpreted it as a stop requested. Excepted this problem of interpretation the precision is correct.

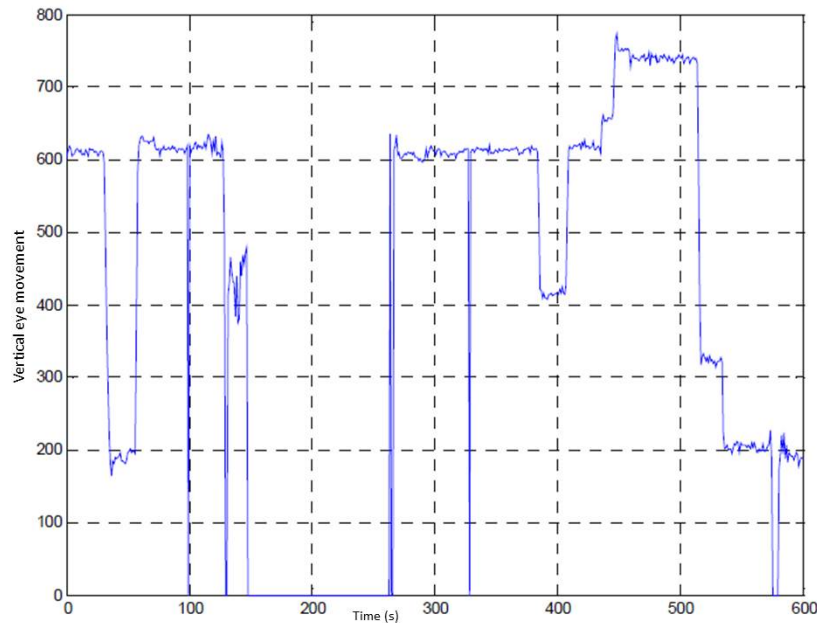


Figure 4 Example of eye movement before alarm.

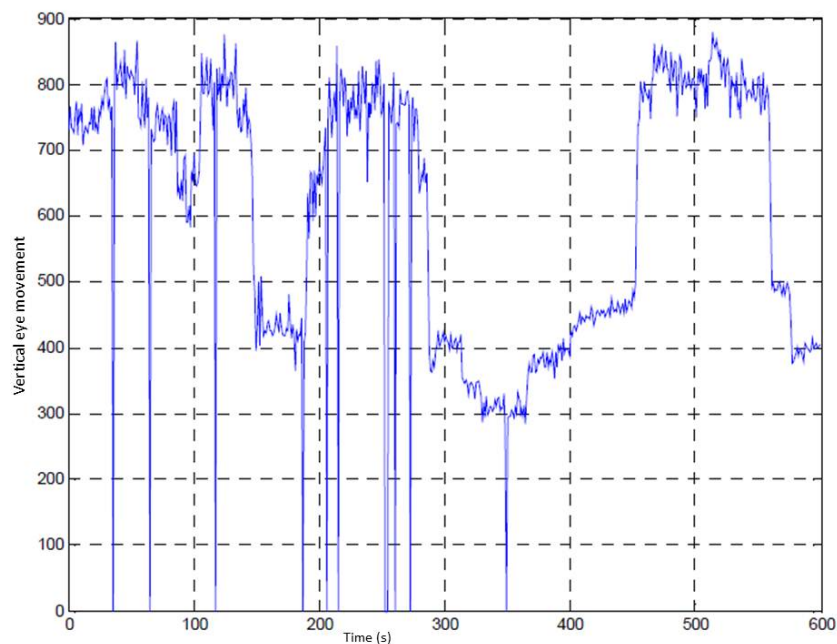


Figure 5 Example of vertical eye movement after alarm.

Figures 4 & 5 illustrate the difference of eye movement of a driver before and after occurrence of an alarm. Before alarm occurrence the movement of eye is relatively “smoothly”. Null value indicates a problem of measurement (eye detection problem, movement of head...).

After the alarm occurrence movements are more “nervous”. The driver watches not only the railway but also the different indicators on the dashboard.

There are some potential differences in terms of behavior between drivers, but also intra-individual differences. For instance, after a disturbance, the eye movement can change. The driver may change his/her strategy to get some information.

The emotional state was detected during this preliminary experiment using Face Reader ® software. Rough results are difficult to interpret. Emotional state variation is very important. Nevertheless, we noticed that frequencies of occurrence of some particular states such as fear, disgusting, and angry, that can be considered as “negative emotions”, were more important after a default occurrence (cf. Figure 6).

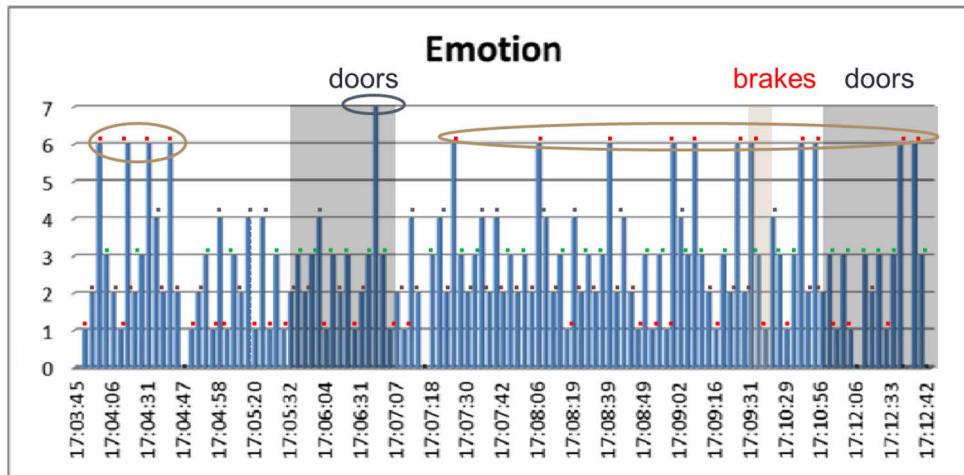


Figure 6 Emotional states before and after default.

Conclusion

This paper presented a preliminary experiment illustrating the interest of non-intrusive systems in order to detect and diagnose the state of the human operator. The combination of different tools, using different technology should improve the diagnosis of the human operator state. The different research teams in the HAMASYT network had some expertise in different non-intrusive tools. For instance, at TU-Delft and University of Valenciennes, competences in facial expression generation are developed. Some eye-tracking systems are used in TU-Berlin and University of Valenciennes in order to analyze vigilance. A specific research topic will be studied for the future HAMASYT network.

Bibliography

- Joost, B., Qu, C. & Brinkman, W-P. (2012). *Dynamic Facial Expression of Emotion Made Easy*. Technical report, Interactive Intelligence, Delft University of Technology.
- Heloir, A. & Kipp, M. (2010). Real-time Animation of Interactive Agents: Specification and Realization. *Applied Artificial Intelligence*. 24(6), 510-529.
- Loslever, P., Simon, P., Rousseau, F. & Popieul, J.C. (2008). Using space windowing for a preliminary analysis of complex time data in human component system studies. Examples with eye-tracking in advertising and car/head movements in driving. *Information Sciences*, 178, 3645 - 3664.

- Onnasch, L., Ruff, S. & Manzey, D. (2012). Operators' Adaption to Unreliability of Alarm Systems: A Performance and Eye-Tracking Analysis. *Proceedings of the 56th Annual Meeting of the Human Factors & Ergonomics Society*. (pp. 1948–1952).
- Popieul, J.-C., Loslever, P., Todoskoff, A., Simon, P. & Roetting, M. (2012). Multivariate analysis of human behavior data using fuzzy windowing: Example with driver-car-environment system. *Engineering Applications of Artificial Intelligence*, 25, 989 - 996.
- Venjakob, A., Marnitz, T., Mahler, J., Sechelman, S. & Roetting, M. (2012). Radiologists' eye gaze when reading cranial CT images. *Proceedings of SPIE Vol. 8318 83180B-1*.
- Wegerich, A., Beckmann, M., Schmuntzsch, U. & Roetting, M. (2013). Reventing Erroneous Operator Behavior and Supporting Optimal Operating Procedures within Industrial Product-Service Systems. *Proceedings of CIRP Conference on Industrial Product-Service Systems*. Bochum, 2013.

Impact of increasing workload on facial temperature in a simulated driving task

Stefan Ruff¹ and Matthias Rötting²

Keywords: workload, facial temperature, non-intrusive, psychophysiology

Abstract

To avoid accidents caused by an inappropriate amount of mental workload it is crucial to continuously and non-intrusively monitor the actual workload state of a driver. Whereas many studies extensively discuss physiological measures like heart rate, heart rate variability and pupil dilation only little research has yet been conducted on facial temperature features. The present study examines the impact of increasing secondary task difficulty on nose-tip and forehead temperature in a simulated driving task. The results implicate that nose-tip temperature differentiates between different workload levels, imposed by a paced mental arithmetic task, whereas forehead temperature does not. The promising results of this experiment need to be validated in future experiments. They promote the integration of facial temperature features in multi-parameter approaches for online workload detection.

Introduction

There is a sustained upward trend in information presented to the driver by infotainment and assistance systems in cars. This may result in a change in demand of attentional resources and could lead to potentially dangerous situations (Brookhuis & de Waard, 2010). For this reason there is a great need for non-intrusive methods to continuously assess the actual workload state of the driver to prevent under- or overload situations.

Mental workload is highly associated with the concept of mental effort (Mulder 1986). With increasing task demands more effort needs to be invested to maintain performance. Therefore a promising approach is to assess sympathetic nervous system (SNS) activation by capturing changes in physiological measures (e.g. O'Donnel & Eggemeier, 1986; Mulder, 1986; Furedy, 1987; Iani, Gopher, Grunwald & Lavie, 2002). Whereas a great number of studies discuss physiological measures like heart rate, heart rate variability and pupil dilation only little research has yet been conducted on facial temperature.

Using skin temperature features could be a promising way to detect changes in mental workload since peripheral blood vessels located under the skin of the nose are solely innervated by the sympathetic nervous system (SNS). The SNS is responsible for vasoconstriction and vasodilation of peripheral blood vessels. In-

¹ Technische Universität Berlin, Institut für Psychologie und Arbeitswissenschaft, FG MMS

² Technische Universität Berlin, Institut für Psychologie und Arbeitswissenschaft, FG MMS

creased activity of the SNS leads to constriction of peripheral blood vessels and hence to a decrease in blood flow. Temperature changes at the tip of the nose should thus be a good indicator of sympathetic activity (Iani, Gopher & Lavie, 2004; Nozawa & Tacano, 2009).

Several studies investigated the impact of different cognitively demanding secondary tasks on facial temperature features and showed that the temperature at the tip of the nose decreases with the introduction of a secondary cognitively demanding task (Veltman & Voß, 2005; Or & Duffy, 2007; Itoh, 2009). Research also showed that forehead temperature remains stable under cognitively or emotionally demanding conditions. This enables the use of the forehead temperature as a reference measure (Genno, Ishikawa, Kanbara, Kikumoto, Fujiwara, Suzuki & Osumi, 1997; Or & Duffy, 2007).

The aim of this study is to examine the impact of three different task difficulty levels on nose-tip and forehead temperature to assess if it is possible to differentiate between different workload levels.

Method

A total of 36 subjects (male=18 and female=18) ranging in age from 20 to 31 ($M = 24.73$, $SD = 2.68$) participated in a dual task paradigm study.

As primary task the participants performed the Lane Change Task (LCT) developed by Mattes (2003) in a fixed base driving simulator at a fixed speed of 80km/h. Parallel to the driving task they performed a modified Paced Auditory Serial Addition Task (PASAT) (Gronwall, 1977) with three degrees of difficulty (low/medium/high, within-subjects). The PASAT was chosen because of its high demands on working memory as well as sustained and divided attention. Participants were asked to continuously sum up the last two auditorily presented digits, ranging from 1 to 9, and to respond verbally in the medium (5 seconds interdigit-interval) and high workload condition (3 seconds interdigit-interval) whereas in the low workload condition (5 seconds interdigit-interval) they were just asked to repeat the digits presented.

For all three conditions the trial consisted of four consecutive two-minute driving sections (pre/p1/p2/post). The first and the last section (pre and post) were baseline sections without a secondary task to take individual differences in initial physiological levels into account. In section two and three the participants performed the modified PASAT in parallel to their driving task (Fig.1).

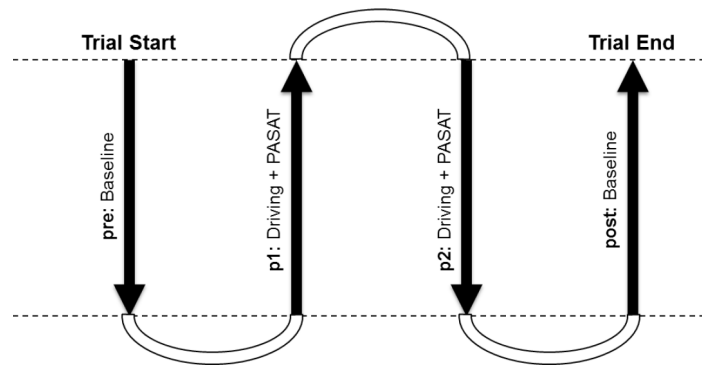


Fig.1: Example trial for one condition

Three different categories of dependent measures were analyzed to cover the different dimensions of mental workload proposed by O'Donnell and Eggemeier (1986): performance, subjective and physiological data. To cope with individual differences in initial performance or physiological measures, we compared changes of the parameters in the dual task condition (p1/p2) with the pre baseline (pre).

(1) Performance measures:

LCT: Primary task performance was operationalized by the lane change accuracy represented by the mean lane deviation from a standardized lane (Fig.2). Primary task performance (Ptp) in percent is calculated as:

$$Ptp = \frac{\text{mean lane deviation}(p1, p2) - \text{mean lane deviation}(\text{pre})}{\text{mean lane deviation}(\text{pre})} * 100\%$$

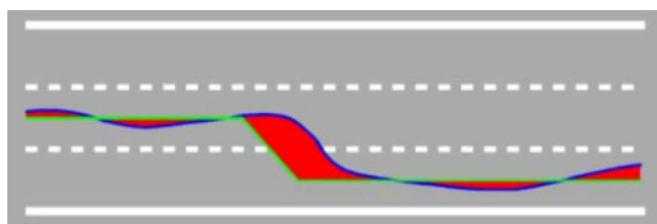


Fig.2: Lane change accuracy as the area (red) between the actual lane change behaviour (blue) and a standardized lane change (green) (Mattes 2003)

PASAT: Secondary task performance (Stp) is calculated as the percentage of errors averaged over p1 and p2 section.

(2) Subjective measures:

Subjective workload was assessed with the NASA-Task Load Index (Hart & Staveland, 1988). Subjective workload ratings are reported as weighted overall score for each trial.

(3) Physiological measures:

Facial temperature features were recorded with a Testo 875-1 thermography camera (optical resolution: 160*120 pixel, emission coefficient: 0.98, thermal resolution: < 0.08K). Two regions of interest were defined: ROI Nose (5x5 pixel) at the tip of the nose and ROI Forehead (10x10 pixel) in the center of the forehead right above the eyebrows (Fig.3)

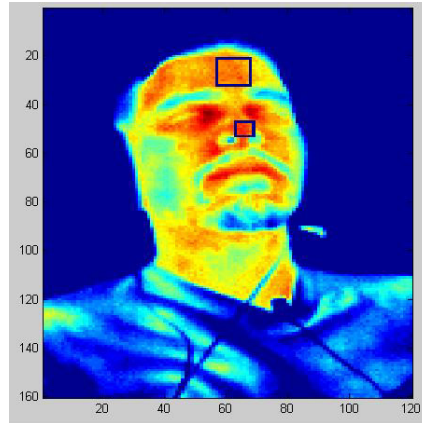


Fig.3: ROI Nose (5x5 pixel) and ROI Forehead (10x10 pixel)

Nose temperature: Nose-tip temperature change (Ntc) is calculated in percent as:

$$Ntc = \frac{\text{mean nosetip temperature}(p1, p2) - \text{mean nosetip temperature}(\text{pre})}{\text{mean nosetip temperature}(\text{pre})} * 100\%$$

Forehead temperature: Forehead temperature change (Ftc) in percent is calculated as:

$$Ftc = \frac{\text{mean forehead temperature}(p1, p2) - \text{mean forehead temperature}(\text{pre})}{\text{mean forehead temperature}(\text{pre})} * 100\%$$

Following an instruction on the LCT and the PASAT, participants were familiarized with the two tasks in a practice session to make sure the instructions were fully understood. Subsequently, they were randomly assigned to one of the six counterbalanced testing sequences. Before each trial the subjects were instructed that their primary task is driving and that they should complete the PASAT as good as possible without disregarding the driving task. After each trial participants were asked to answer the NASA-TLX (Hart & Staveland, 1988). Subsequently they were given a period of rest (3 min) before the next trial started.

Results

Six participants had to be removed from data analysis due to missing data. Statistical analysis was done using IBM SPSS Statistics 21. For normally distributed data repeated measures ANOVAs were used. Post-Hoc main effect comparisons were conducted using the Bonferroni correction. For non-normally distributed data related-samples Friedman's Two-Way Analysis of Variance by Ranks was performed.

(1) Performance measures:

Primary task performance (Ptp) slightly decreased with increasing task difficulty (Tab.1). However, a repeated measures ANOVA with three levels (Task difficulty) revealed only a marginally significant effect for task difficulty, $F(2, 58) = 2.878$, $p = 0.064$.

For secondary task performance (Stp) there was a clear impact of task difficulty (Tab.1). In the low workload condition participants produced no errors, whereas for the more difficult conditions we found an increase in the error rate (Fig.4a). This was statistically supported by a highly significant task difficulty effect, $\chi^2(2)=41.543$, $p < 0.001$. Pairwise comparisons revealed statistical significance for the difference between low and medium ($p < 0.001$), low and high ($p < 0.001$) as well as medium and high ($p < 0.05$) workload conditions.

Tab.1: Mean values and standard deviations for primary- and secondary task performance

	Low Workload	Medium Workload	High Workload
Ptp	M= -1.07 (SD= 6.10)	M= 0.21 (SD= 9.05)	M= 3.41 (SD= 8.59)
Stp	M= 0.00 (SD= 0.00)	M= 5.00 (SD= 6.12)	M= 11.61 (SD= 10.24)

(2) Subjective measures

The weighted overall score of perceived workload by the participants showed the expected pattern (Fig.4b). The workload ratings increased as a function of task difficulty with lowest ratings for the low workload condition (low $M = 27.94$, $SD = 14.21$, medium $M = 52.32$, $SD = 15.80$ and high $M = 64.88$, $SD = 17.52$). A repeated measures ANOVA revealed a highly significant effect for task difficulty on experienced workload, $F(2,58) = 72.550$, $p < 0.001$. A post-hoc comparison of the main effect with Bonferroni correction showed that overall workload ratings differentiate between all three conditions ($p < 0.001$).

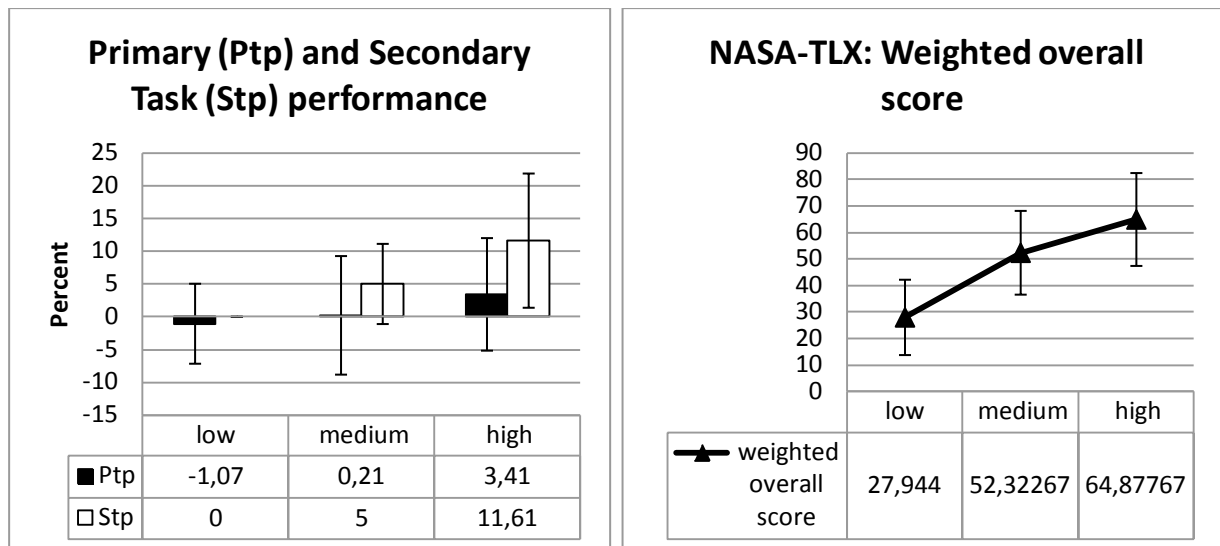


Fig.4: (a) Effect of task difficulty on primary- and secondary task performance (b) Effect of task difficulty on NASA-TLX weighted overall score

(3) Physiological measures

For the forehead temperature change (Ftc) there was only minimal change over all three task difficulty levels (Tab.2). A related-samples Friedman's Two-Way Analysis of Variance by Ranks revealed no significant difference in forehead temperature change (Ftc) for the three task difficulties ($p = .905$).

Tab.2: Mean values and standard deviations for forehead- and nose-tip temperature

	Low Workload	Medium Workload	High Workload
Ftc	M= 0.00 (SD= 0.25)	M= 0.04 (SD= 0.36)	M= -0.05 (SD= 0.17)
Ntc	M= -0.66 (SD= 1.27)	M= -1.58 (SD= 1.58)	M= -1.98 (SD= 1.72)

Regarding the nose-tip temperature change (Ntc) there is a clear trend for decreasing nose-tip temperatures with increasing task difficulty (Fig.5). This is supported by a highly significant main effect, $\chi^2(2)=18.600$, $p < 0.001$. Pairwise comparisons showed a significant difference between low and both medium and high workload conditions ($p < .05$), but no difference between medium and high workload conditions ($p = .364$).

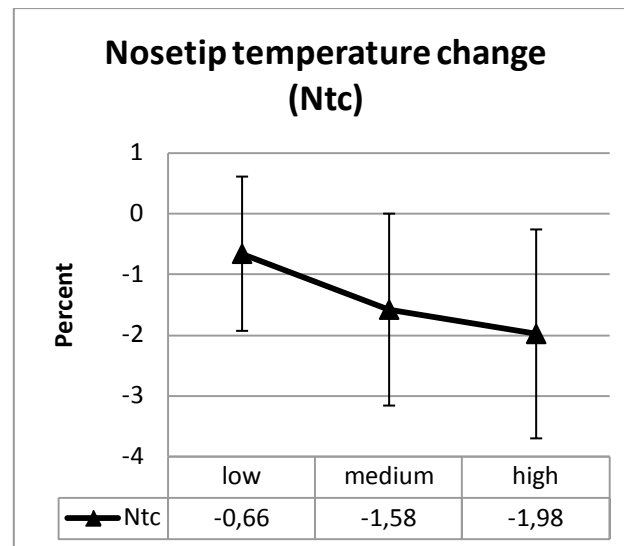


Fig.5: Effect of task difficulty on nose-tip temperature change.

Conclusion

The results indicate that the temperature at the tip of the nose differentiates adequately between different workload levels. Furthermore, the stable forehead temperature confirms the approach of Or and Duffy (2007) to use it as a reference measure. In summary, it can be stated that facial temperature parameters could be a supplement in multi-parameter workload detection approaches. However, it seems not sensitive enough to distinguish between different time demands in a cognitive task. More studies on nose-tip temperature should examine the effect of different cognitive secondary tasks. According to Iani et al. (2004) tasks involving a large amount of central processing resources seem to be associated with strong sympathetic activity whereas tasks that demand high response requirements are reflected in strong parasympathetic activity. Therefor nose-tip temperature changes should be more sensitive to central processing demands than to response demands. As single workload measure, nose-tip temperature only identifies specific workload changes. For a more complex workload detection involving both parasympathetic and sympathetic effects of cognitively demanding tasks, future research should combine nose-tip temperature with other physiological measures like heart rate or pupil dilation.

Literature

- Brookhuis, K. A., & Waard, D. de. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident; analysis and prevention*, 42(3), 898–903. doi:10.1016/j.aap.2009.06.001
- Furedy, J. J. (1987). Beyond heart rate in the cardiac psychophysiology assessment of mental effort: the T-wave amplitude component of the electrocardiogram. *Human Factors*, 29(2), 183-194

- Genno, H., Ishikawa, K., Kanbara, O., Kikumoto, M., Fujiwara, Y., Suzuki, R., & Osumi, M. (1997). Using facial skin temperature to objectively evaluate sensations. *International Journal of Industrial Ergonomics*, 19(2), 161–171. doi:10.1016/S0169-8141(96)00011-X
- Gronwall, D. M. A. (1977). Paced Auditory Serial-Addition Task: A measure of recovery from concussion. *Perceptual and Motor Skills*, 44(2), 367–373. doi:10.2466/pms.1977.44.2.367
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology* (pp. 139–183). Elsevier.
- Iani, C., Gopher, D., Grunwald, A. J., & Lavie, P. (2002). Peripheral Arterial Tone as an On-Line Measure of Flight Load. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46(1), 140–144. doi:10.1177/154193120204600129
- Iani, C., Gopher, D., & Lavie, P. (2004). Effects of task difficulty and invested mental effort on peripheral vasoconstriction. *Psychophysiology*, 41(5), 789–798. doi:10.1111/j.1469-8986.2004.00200.x
- Itoh, M. (2009). Individual differences in effects of secondary cognitive activity during driving on temperature at the nose tip. In *Mechatronics and Automation, 2009. ICMA 2009. International Conference on* (pp. 7–11). Retrieved from <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5246188>
- Mattes, S. (2003). The Lane Change Task as a tool for driver distraction evaluation. Retrieved from <http://people.usd.edu/~schieber/docs/MATTES2003-powerpoint.pdf>
- Mulder, G. (1986). The concept and measurement of mental effort. In G.R.J. Hockey, A.W.K. Gaillard & M.G.H. Coles (Eds.), *Energetics and Human Information Processing* (pp. 175–198). Dordrecht, The Netherlands: Martinus Nijhoff Publisher
- Nozawa, A. & Tacano, M. (2009). Correlation analysis on alpha attenuation and nasal skin temperature. Retrieved from http://iopscience.iop.org/1742-5468/2009/01/P01007/pdf/1742-5468_2009_01_P01007.pdf
- O'Donnell, R. D., and Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman & J. Thomas (Eds.), *Handbook of perception and human performance: Volume II. Cognitive processes and performance* (pp. 42/1-42/49). New York: Wiley.
- Or, C. K. & Duffy, V. G. (2007). Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement. *Occupational Ergonomics*, 7, 83-94.
- Veltman, H. & Vos, W. (2005). Facial Temperature as a Measure of Operator State. In: D. Schmorrow (Hrsg.), *Foundations of augmented cognition* (S. 293 - 301). Mahwah, USA: Lawrence Erlbaum Associates, Inc.

A support tool for assisting human diagnoses of command errors detected by a robust filter

Serge Debernard¹, Pascale Marangé², François Gellot³, Marie-Pierre Pacaux-Lemoine¹, Thierry Poulain¹, Alexandre Philippot³, Bernard Riera³ and Jean-François Pétin²

Keywords: Manufacturing system, Robust filter, Support tool, Human diagnosis.

Abstract

In manufacturing systems, one way to limit serious consequences of an unsuitable command is to detect dangerous forbidden states of the process and to stop it. That can be performed by a robust filter placed inside a PLC, independently of the command. This approach maintains process safety by avoiding damage to products and manufacturing system. But the filter's decision consisting in stopping process may suppress important information that human operators can use to diagnose and correct the command. Therefore, the implementation of a filter must be complemented with some functionality that can assist human operators to diagnose command errors.

This study called ADEXEC (Approach to Detect and EXplain Errors of Command) took place within the framework of the GIS 3SGS project (French acronym for Scientific Interest Group – Systems Monitoring, Dependability and Security) and was realized with partners from the University of Reims (CReSTIC), Nancy (CRAN UMR CNRS 7039) and Valenciennes (LAMIH UMR CNRS 8201).

Introduction

Ensuring dependability of manufacturing system control is currently based on two complementary approaches. The first one focuses on the control design activities with the objective to avoid unexpected behavior. Two main approaches are suggested in this way (Faure J-M. & al., 2001): (i) control validation and verification (Berard & al., 1999 ; Roussel & al., 2002), and (ii) Supervisory Control Theory based on synthesis controller (Ramadge & al., 1989). Most of the time, those designing approaches make two strong assumptions: the behavior of plant devices is not faulty and the designed control is exactly the same as the program that is implemented on the control devices (i.e. code generation deviations or code modifications by maintenance agents are not considered).

¹ Laboratory of Industrial and Human Automation, Mechanics and Computer Science, LAMIH, CNRS UMR 8201, University of Valenciennes, Le Mont Houy, F-59313 Valenciennes Cedex 9, France

² Centre de Recherche en Automatique de Nancy, CRAN, CNRS UMR 7039, University of Lorraine, Faculté des Sciences, BP 70239, F-54506 VANDOEUVRE LES NANCY Cedex, France

³ CReSTIC URCA, UFR Sciences Exactes et Naturelles, Moulin de la Housse - BP 1039, F-51687 REIMS Cedex 2, France

These assumptions being not realistic in practice, the second approach complements the safe design activities by the development of online barriers that enable the monitoring, the diagnosis or even the reconfiguration of the plant. More precisely, diagnosis of manufacturing systems aims at detecting unsafe behavior of the plant and localizing the components that are involved in the behavioral deviation while control filtering aims at avoiding that a PLC program provokes plant damages, whatever the PLC program (Marangé, 2008), (Riera & al., 2012). The filter is placed between the controller and the plant and inhibits potential dangerous evolutions by checking a set of safety constraints. At least, if the plant situation is unknown, automatic procedures implemented by control filtering and diagnosis may be not efficient. This case generally requires the intervention of expert to analyze the unknown situation of the plant, and to take emergency decision to drive back the plant in acceptable states.

All these complementary mechanisms (diagnosis, control filtering, human/system cooperation) are generally studied independently from each other. Trying to define interactions between all these online safety mechanisms and proposing a consistent engineering framework for the design of these barriers are the objectives of the ADEXEC project (Figure 1).

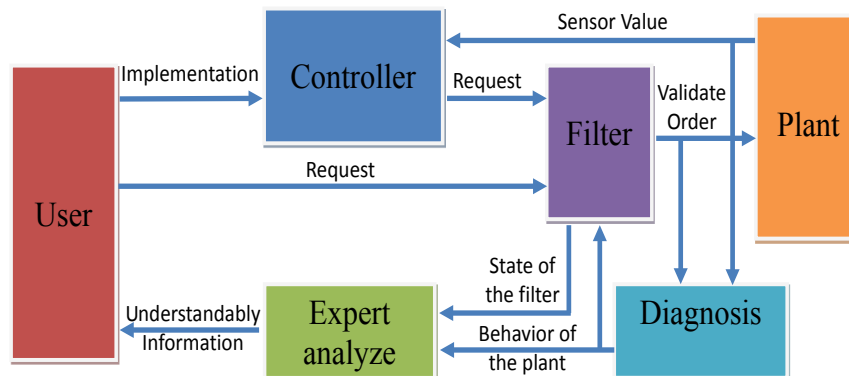


Figure 1: Principle of the project

This project addresses the following scientific problems:

- The constraints that are applied within the control filter assume that all information coming from the plant is reliable; consequently, in case of a device fault, the filter may inhibit actions that should be safe or, worse, authorize actions that should be forbidden: taking in account diagnosis information when checking the constraint should tackle this drawback,
- Diagnosis is built by taking into account the designed control rules and without taking into account filtering actions that correct the deviations of the control system: taking into account the information about satisfactory of the filters constraints should help in detecting and localizing plant faults
- At least, the generation of the explanations about the current plant situation to a human operator must be understandable and enough synthetic to be compatible with his level of knowledge. Providing automatic generation of dif-

ferent levels of alarms from the satisfactory of the filter constraints should help the human operator to better handle the real plant state.

The paper is organized as follows. Section 2 presents the works on the definition of robust filter. Section 3 presents a support tool for assisting human diagnoses. Sections 4 and 5 present the experimentation and the first results on the case study: sorting process. Section 6 concludes the paper.

Robust filter

The proposed methodology to design the filter is based on the use of safety constraints. This filter is obtained from a modular and iterative approach of modeling and validation in order to avoid complexity and combinatory explosion and to ensure its applicability in industrial world (Benlorhfar & al., 2011; Marangé & al., 2010). The necessary 4 steps are presented and briefly described in the following points:

- Expert brings a dysfunctional analysis to define all dangerous states, which will never be reached by the elements of plant or products. For each dangerous state, a set of properties is defined and formally checked.
- For each dangerous state, expert identifies the subset of the plant (system components and products flows sequence) to model for the validation by model-checking.
- For each subset, expert defines the set of safety constraints, which will compose the filter. He/she also identifies the necessary observers (reconstruction of the missing information, unobservability).
- Expert validates the set of constraints for all subsystems by checking the properties.

In the presented approach, it is assumed that the safety constraints can always be represented as a monomial and depend on the inputs E, outputs S (current time (t) and previous times (t-1, t-2...)) and observers O (depending only inputs at the current time (t) and previous times (t-1, t-2...)).

$Sk = Sk(t)$: logical variable corresponding to the value of PLC boolean output at time t and S^* : set of variables Sk at times t-1, t-2 ...

Two types of constraints are defined: simple safety constraints (SCs, equation (1)) and the combined safety constraints (SCc, equation (2)). The combined constraints involve only two outputs.

$$SC_{si}(Sk, Y) = h_i(Sk, E, O, S^*) \quad (1)$$

$$SC_{ci}(Sk, Sl, Y) = h_i(Sk, Sl, E, O, S^*) \quad (2)$$

Simple safety constraints (SCs) can be written in two different ways (equations (3) or (4)).

$$SC_{si}(Sk, Y) = Sk. f_i(E, O, S^*) \quad (3)$$

$$SC_{si}(Sk, Y) = \overline{Sk} \cdot f_i(E, O, S^*) \quad (4)$$

These simple safety constraints express the fact that if $f_i(E, O, S^*)$ is equal to 1, Sk must be necessarily equal to 0 ($Sk \cdot f_i(E, O, S^*)$) or equal to $1\overline{Sk} \cdot f_i(E, O, S^*)$).

The combined safety constraints (SC_c) are expressed as monomial and involve several outputs Sk, Sl, \dots (equation 5).

$$\sum_j SC_{c_j} = \sum f_{kl\dots}(E, O, Sk, Sl, \dots) = 0 \quad (5)$$

$f_{kl\dots}(E, O, Sk, Sl, \dots)$ is a logic function only dependent of Sk, Sl, \dots at t . In most cases, these constraints depend only on two variables.

A support tool for assisting human diagnoses

According to Ribert-Van De Weerd & al. (2000), maintenance activity that supports the detection of component failures is split into three steps: the first is the preparation of the intervention, that is, the organization of the intervention on the process according to a pre-diagnosis. The second is the intervention and includes diagnosis and carrying out of the work. Finally, the last phase occurs after intervention. According to the authors, the second phase consists of several stages: i) The acquisition of clues and their categorization, ii) The refinement of this categorization and, iii) The gradual reduction of assumptions and the localization of the faulty component, and iv) the validation step.

According to Rasmussen (1986), the cognitive processes of maintenance workers are directed by the normal and abnormal functioning of the system, and searches can be functional (functions performed by a component or subsystem) or topographical. In our study, the filter stops the operating part of process as soon as a dangerous forbidden state is detected. So, the maintenance agent has no clues to direct her/his search and to detect the error in the command. Fig. 2 represents such a situation.

Four levels of explanation have been defined in order to assess the adaptation of the level of explanation according to the type of human operator who would have to diagnose. But these levels have been also defined in order to assess the pertinence of our propositions. Explanations provide information inferred from the logical constraints and process state. The higher is the level of explanation, the more the information is “intelligent”, but the more the explanation is difficult to extract for the designer of the filter and to understand for the human operator using this system. Four levels of explanation have been represented on a top view of the process. It represents sensors and actuators, and information according to the levels of explanation. This kind of representation supports topographic searches of clues. The same top view is also used to present the “current view”, displaying sensors and actuators states at the stopped time (Fig. 3). First level of explanation is proposed by the “projection view”, that is, what would have been the state of the process if the process had not been stopped by the assistance system (Fig. 4). This is

in fact the representation of the bad consequences due to command error; it is a functional representation of the problem.



Fig 2. Sorting process of cases

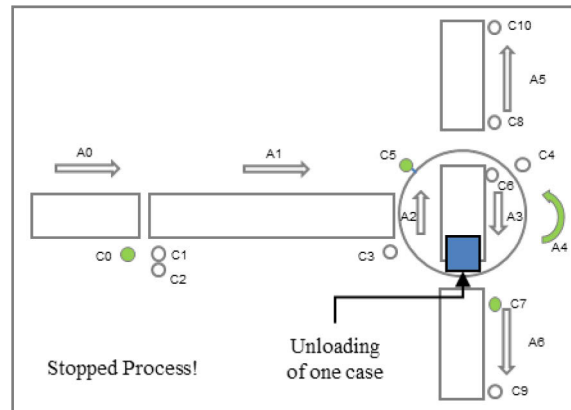
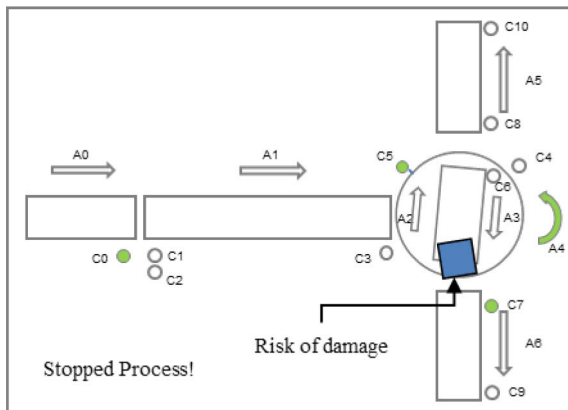
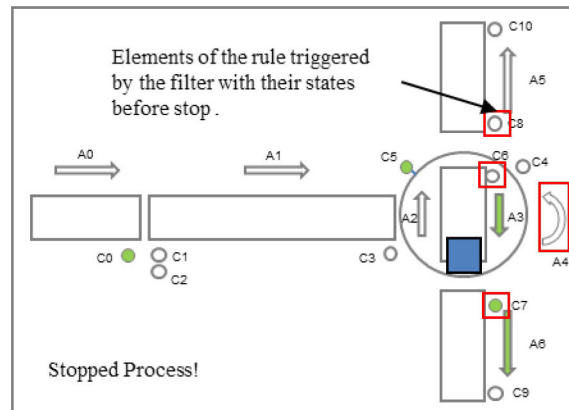


Fig. 3. “Current” view



**Fig. 4. First level of assistance:
« Projection » view**



**Fig. 5. Second level of assistance:
« Status before stop » view**

The second level of explanation, “status before stop”, is the presentation of the state of the process just before the stop (Fig. 5). This representation focuses on the process elements that had triggered the filter and presents the result of the erroneous command. The third level of explanation, “advanced representation”, is the presentation of the events (rising and falling edge) which had reached a forbidden process state (Fig. 6). Here, the display presents two process states, before the forbidden state and after, and explains the bad behavior of the command.

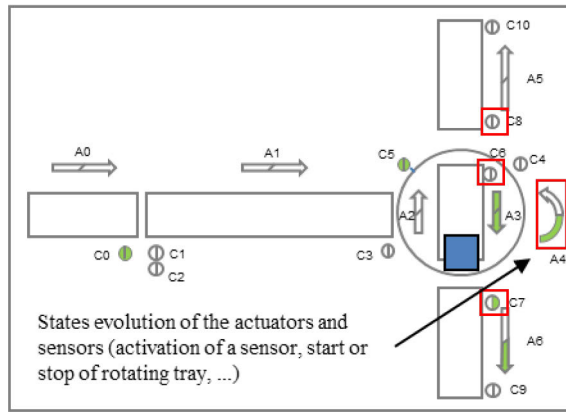
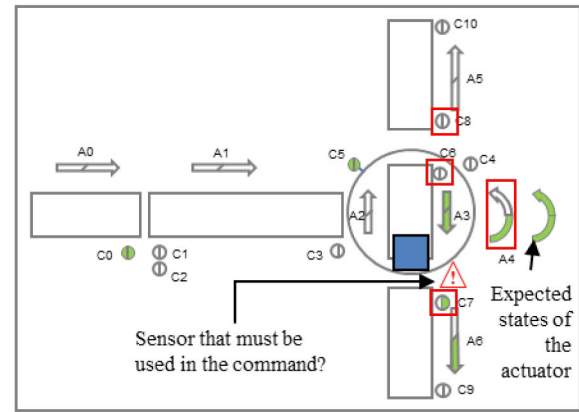


Fig. 6. Third level of assistance:
« Advanced representation » view



7. Fourth level of assistance: « Hy-
pothesis on the command » view

The fourth level of explanation is the presentation of hypotheses concerning the command errors (Fig. 7). This representation explains the behavior that the command should have, and presents information concerning the non-use of information to reach an output, or the explanation of an unwanted output.

Experimentations

An experimental protocol has been designed in order to evaluate these levels of explanation. Each participant is tested individually during two hours at most. On arrival, participants are briefed on the project objectives and experimental protocol. They read and sign an informed consent document and fill in a brief demographic questionnaire in order to be informed about their know how regarding the automation domain (knowledge-based). Next, they have a short training with the experimental platform to improve their know-how regarding process type (rule or skill-based). They learn to control a sorting process. The operative process is simulated on a computer connected to a PLC. Participants are also trained with the assistance system; they learn to use all the levels of explanation provided by the support tool.

Participants performed three tests. They received a photo of the stopped process (Fig. 2) and the “current view” of the process for each test (Fig. 3). The first test is the reference condition, and participants received no explanation and no information concerning products. The second test deals with the first level of explanation provided by the “projection view” (Fig. 4). The last test is one of three more advanced levels of assistance provided by the view corresponding to the level. Participants tested only one advanced level of assistance in order to avoid learning effects regarding the use of assistance system. For the same reason, the command and the type of command error are different for each test. Scenarios and levels of assistance are crossed in order to avoid carryover effects.

Three types of measurements have been done. Participants filled in questionnaires regarding their perception of the assistance system. They had to choose an answer between 1 and 7 to assess readability, comprehensibility, and relevance of the explanations provided by the assistance system. Secondly, participants’ situa-

tion awareness has been assessed regarding the operative part of the process and the information provided by the assistance system. Thirdly, error detection and error correction performance have been assessed.

First Results

The experiments have been conducted with 9 participants. The sample was composed of 9 men and 0 women, with a mean age of 22.11 years, ranging from 21 to 27 years ($SD = 1.96$ years). The results from subjective data are stemmed from questionnaire filling. Concerning their own evaluation of their level in automation, GRAFCET and Ladder, they are relatively confident in their know-how (answer between 1 and 7; Automation: $M: 4.77$ ($SD = 0.67$); GRAFCET: $M: 5.55$ ($SD = 0.88$); Ladder: $M: 5.44$ ($SD = 0.72$)). The evaluation of experiments by participants is quite coherent with previous results; Participants assessed that they are enough trained with the process we had proposed ($M: 5.66$ ($SD: 1.87$)), with assistance system ($M: 5.44$ ($SD: 1.01$)). But they assessed that command errors detection are difficult (Easy to find, answer between 1 and 7; $M: 3.66$ ($SD: 1.41$)). Nevertheless they assessed that their level in automation is relatively sufficient for detecting and correcting command errors ($M: 5.55$ ($SD: 0.58$)).

Subjective data: Results stemmed from subjective data are the evaluation of assistance systems by participants. They assessed two types of assistance systems, the “projection” view (PV) (Fig. 4) and one more advanced view (AV) (Fig. 5, 6 or 7). They assessed that information proposed by assistance systems are readable (PV: $M: 6.33$ ($SD: 0.70$)); AV: $M: 6.11$ ($SD: 1.05$)), comprehensible (PV: $M: 5.77$ ($SD: 0.97$)); AV: $M: 6.11$ ($SD: 0.92$)), explanations are relevant (PV: $M: 5.44$ ($SD: 1.23$)); AV: $M: 5.77$ ($SD: 0.83$)), analysis are comprehensible (PV: $M: 5.83$ ($SD: 0.86$)); AV: $M: 5.44$ ($SD: 1.50$)), assistance systems help in detecting command errors (PV: $M: 5.33$ ($SD: 1.87$)); AV: $M: 6.22$ ($SD: 0.83$)) and in correcting commands (PV: $M: 5.66$ ($SD: 0.86$); AV: $M: 6.0$ ($SD: 1.0$)).

Objective data: Results stemmed from objective data underline that performance is better with the more advanced view. Respectively, with only “current view” (CV), “projection” view (PV) and more advanced view (AV), participants found:

- the part of GRAFCET responsible for command errors: CV: $M: 0.44$; PV: $M: 1.0$; AV: $M: 1.0$,
- the specific step in the GRAFCET: CV: $M: 0.33$; PV: $M: 0.55$; AV: $M: 0.88$,
- the specific ladder: CV: $M: 0.22$; PV: $M: 0.38$; AV: $M: 0.66$,
- the explanation of their correction: CV: $M: 0.33$; PV: $M: 0.77$; AV: $M: 0.72$.

Participants fulfilled the situation awareness questionnaire; they indicated what they kept in memory concerning process state and information provided by assistance system. It is the first level of situation awareness proposed by Endsley and al. (1983). They are trained to do that and it seems to be a bias in the experiment, be-

cause they forced themselves to recall all information even if didn't need them to understand process state.

To sum up, the general trend is that the more the assistance system provides explanations about its analysis of the process, i.e. if participants have more explanations than just the current and projection views, the more they detect and correct the error in the command.

Conclusion

This paper has presented shortly the ADEXEC project. In this project, a robust filter is used in order to avoid product damages or plant damages in the case of bad code implemented in PLC for example, or bad command sent by human operator. But when the robust filter detects a dangerous forbidden state, the process is stopped and the maintenance agent has no clues for performing a diagnosis and to find the error in the PLC program. Based on the concept of human-machine cooperation, some new IHMs had been proposed in order to help the human operator and some experiments had been realized. The first results of these experiments had been presented and showed that the more the assistance system provides explanations about its analysis of the process, the more human agents detect and correct the error in the command. Nevertheless, it is necessary to perform more experiments in order to validate these first results.

Literature

- Benlorhfar, R., Annebicque, D., Gellot, F. & Riera, B. (2011). Robust filtering of PLC program for automated systems of production, 18th World Congress of the International Federation of Automatic Control, Milano, Italy, august.
- Berard B., Bidoit M., Finkel A., Laroussinie F., Petit A., Petrucci L. & Schnoebelen P. (1999). *Systems and software verification: Modelchecking techniques and tools*, Heidelberg, Springer-Verlag Edition.
- Endsley, M. & Robertson M. (1983). Situation awareness in aircraft maintenance teams. *International Journal of Industrial Ergonomics*, 26, pp 301-325. 5 779.
- Faure J-M. & Lesage J-J.(2001), Methods for safe control systems design and implementations, 10th IFAC Symposium on Information Control Problems in Manufacturing, INCOM'2001, Vienna, Austria.
- Marangé P., Benlorhfar R., Gellot F. & Riera B. (2010). Prevention of human control errors by robust filter for manufacturing system, 11th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Valenciennes, France.
- Marangé P. (2008). Synthèse et filtrage robuste de la commande pour des systèmes manufacturiers surs de fonctionnement, Thèse de doctorat, Université de Reims Champagne-Ardenne.

- Ramadge G. & Wonham W. M. (1989). The control of discrete event systems, Proc. IEEE, Special issue on DEDSs, 77, pp.81-98.
- Rasmussen J. (1986). Information processing and human-machine interaction. Amsterdam: Elsevier Science.
- Ribert-Van De Weerd C. & Brangier E. (2000). L'usage et l'efficacité des aides à la maintenance en télédiffusion. *Le travail Humain*, 63(4), pp 331-352, Presse Universitaire de France.
- Riera B., Annebicque D., Gellot F., Philippot A. & Benlorhfar R. (2012). *Control synthesis based on logical constraints for safe manufacturing systems*, Bucharest, Roumanie.
- Roussel J.M. & Faure J.M. (2002). An algebraic approach for PLC programs verification, In *Proceedings of 6th international Workshop On Discrete Event Systems*, Zaragoza, Spain, pp. 303–308.

Driver Distraction: Weiterentwicklung einer Methode zur Messung von kognitiver Beanspruchung durch Anzeige-und Bedienkonzepte im Fahrzeug

Odette Mantzke¹, Roman Vilimek¹ und Claus Christian Carbon²

Schlüsselwörter: Cognitive Workload, Detection Response Task, Pupillometrie, komplexe Fahrszenen

Zusammenfassung

Die Messung von kognitiver Beanspruchung (Workload), welcher durch Bedienung von Fahrerassistenz- und Informationssystemen entsteht - wird zunehmend wichtiger. Dazu werden valide Messmethoden benötigt. In einer Fahrsimulatorstudie mit 24 Probanden wurden verschieden komplexe Fahrszenarien hinsichtlich des entstandenen Workload bewertet, sowohl unter Bearbeitung einer visuell-manuellen (SuRT) als auch einer kognitiv beanspruchenden Nebenaufgabe (n-back-task). Workload wurde hierbei mittels einer Detection Response Task (DRT) sowie des Index of Cognitive Activity (ICA) gemessen. Die Reaktionszeiten der DRT unterschieden sich signifikant zwischen einer nicht komplexen Landstraßenfahrt und einer komplexen Stadtfahrt. Außerdem konnten signifikante Reaktionszeitunterschiede in Bezug auf die Schwierigkeitsstufen der n-back-task nachgewiesen werden. Der ICA konnte diese Unterschiede nicht messen. Somit konnte gezeigt werden, dass die DRT ein robustes Instrument zur Messung von Workload im Fahrzeugbereich darstellt, welches in komplexen Fahrsituationen Beanspruchung zuverlässig messen kann. Hinsichtlich des ICA scheinen hingegen weitere Untersuchungen nötig zu sein, um die Anwendbarkeit in einem anwendungsnahen Setting ermitteln zu können.

Einleitung

Automobilhersteller tragen die Verantwortung dafür, dass Funktionen wie Fahrerassistenz- und Informationssysteme (IVIS) während der Fahrt sicher bedienbar sind. Ein mögliches Sicherheitsrisiko besteht in der Ablenkung, die bei Bedienung von IVIS auftreten kann. Bisherige Verfahren zur Messung dieser Ablenkung wie z. B. die Maßgaben der Alliance of Automobile Manufacturers (AAM, 2006) geben für die Bewertung des sicheren Fahrens Grenzwerte in Fahr- und Blickdaten vor. Der Nachteil an dieser Vorgehensweise besteht darin, dass sie sich nur auf die Messung visuell-manueller Ablenkung konzentriert. Die zu Grunde liegende kognitive Beanspruchung wird nicht berücksichtigt. In Zukunft werden IVIS - z.B. durch Sprachbedienung - jedoch vermehrt kognitiven statt visuell-manuellen

¹ BMW Group, Anzeige- und Bedienkonzepte, München

² Otto-Friedrich-Universität Bamberg, Lehrstuhl für Allgemeine Psychologie

Workload verursachen. Im Rahmen eines Forschungsprojektes der BMW Group soll zu diesem Zweck eine Methode zur Messung von kognitiver Beanspruchung entwickelt werden, mit der ein mögliches Sicherheitsrisiko von IVIS und Auswirkungen auf das Fahrverhalten valide bewertet werden können. Zur adäquaten Bestimmung von Workload werden dabei Methoden benötigt, die Workload kontinuierlich und mit einer hohen zeitlichen Auflösung messen. Mögliche Kandidaten dafür sind die DRT und der ICA (Vilimek, Schäfer & Keinath, 2012; Schwalm, 2009). Im Fahrsimulator wurden beide Methoden meist bei einfachen Fahraufgaben eingesetzt (z.B. Conti, Dlugosch, Vilimek, Keinath & Bengler, 2012). Um bei der Bestimmung von Workload möglichst realistische Aussagen treffen zu können, sollte untersucht werden, ob diese Methoden auch in komplexeren Fahrszenarien Workload zuverlässig messen. Der Fokus der vorliegenden Untersuchung liegt daher in der Anwendbarkeit dieser beiden Methoden in komplexeren Fahrszenarien unter Bearbeitung von Fahr- und Nebenaufgabe.

Die DRT ist eine Reiz-Reaktionsaufgabe, bei der Workload über einen Anstieg der Reaktionszeit auf einen vorgegebenen Reiz gemessen wird. Die Methode bietet den Vorteil, dass sie zum einen sehr sensitiv gegenüber Workloadveränderungen (Jahn, Oehme, Krems & Gelau, 2005) und zum anderen sehr einfach und ökonomisch in der Anwendung und Auswertung ist. Ein Nachteil der DRT besteht darin, dass sie als eigenständig zu bearbeitende Aufgabe Workload zusätzlich erhöht und bei Anwendung auf Fahr- und Nebenaufgaben ein triple-task-setting erschafft. Der ICA bedient sich zur Workloadmessung kurzfristigen und schnellen Änderungen der Pupillengröße, sogenannten "bursts of dilations", welche bei Beanspruchung vermehrt auftreten (Schwalm, Keinath, & Zimmer, 2008). Die Messung des ICA erfolgt über einen Eyetracker und hat nicht das Problem des triple-task-settings. Sie ist jedoch sehr aufwendig in der Anwendung und kann zudem nur über einen nicht offengelegten Algorithmus ausgewertet werden. Beide Methoden sollen hinsichtlich der Anwendbarkeit in unterschiedlich komplexen Fahrszenarien miteinander verglichen werden. Es werden daher folgende Hypothesen untersucht:

- Es ist bekannt, dass die Komplexität einer Fahrszene einen Einfluss auf den Workload des Fahrers hat (Jahn et al., 2005). Der Anstieg des Workloads von einfachen zu komplexen Fahrszenarien sollte sich in einer Fahrt ohne Nebenaufgabe sowohl in einem Anstieg der Reaktionszeit auf die DRT als auch im steigenden ICA-Wert zeigen.
- Der Workload ist während der Bearbeitung einer Nebenaufgabe höher als in der Baseline-Bedingung. Außerdem erzeugen komplexe kognitive Nebenaufgaben einen höheren Workload als leichte kognitive Nebenaufgaben. Beide Messinstrumente (ICA und DRT) sollten diesen Anstieg zeigen.

Methode

An der Studie nahmen 21 Männer und 3 Frauen teil. Die Probanden waren Mitarbeiter der BMW Group und zwischen 21 und 51 Jahre alt ($M = 34.8$ Jahre, $SD = 7.7$ Jahre). Die Untersuchung fand in einem statischen Fahrsimulator im Zentrum für Fahrsimulation der BMW Group in München statt. Die Strecke enthielt drei Abschnitte, welche sich hinsichtlich ihrer Komplexität unterschieden. Bei der Konstruktion der Streckenabschnitte wurde sich an der Klassifizierung von Fastenmeier (1995a) und Rösler (2010) orientiert, sodass sich die Abschnitte in der Anzahl der fahrrelevanten Reize sowie in der Stärke des Handlungseingriffs unterschieden. Die Strecke bestand aus einem kurvenreichen Landstraßenabschnitt mit wenig Fremdverkehr, einem Stadtabschnitt mit wenig Fremdverkehr sowie einem komplexen Stadtabschnitt mit viel Fremdverkehr sowie vorfahrtgewährenden Kreuzungen. In jedem Abschnitt bearbeiteten die Probanden mehrere Wiederholungen einer kognitiv beanspruchenden Nebenaufgabe und einer visuell-manuell beanspruchenden Nebenaufgabe in zwei Schwierigkeitsstufen für je zwei Minuten. Zudem enthielt jeder Abschnitt eine zweiminütige Baseline-Fahrt, in der die Probanden keine Nebenaufgabe bearbeiteten. Jeder Abschnitt hatte somit eine Dauer von 10 Minuten, der komplette Versuch dauerte 1,5 Stunden. Als kognitiv beanspruchende Nebenaufgabe wurde die n-back-task nach Mehler, Reimer, Coughlin & Dusek (2009) in den Ausprägungen 0-back und 2-back verwendet. Als visuell-manuell beanspruchende Aufgabe bearbeiteten die Probanden den Surrogate-Reference task (SuRT) in den in der ISO Norm 14198 vorgegebenen Schwierigkeitsstufen leicht und schwer. Zur kontinuierlichen Workloadmessung wurde zum einen eine head-mounted Detection Response Task (DRT) verwendet. Hierfür wurde eine LED an einem Basecap befestigt, welche im Abstand von 3-6 Sekunden rot leuchtete. Mittels eines Tasters am Lenkrad konnten die Probanden die Wahrnehmung des Reizes bestätigen. Zum anderen wurde der Index of Cognitive Activity (ICA) verwendet. Der Pupillendurchmesser wurde mittels eines EyeLink II Systems der Firma SR Research aufgenommen. Die Probanden übten zunächst alle Nebenaufgaben sowie die DRT im Stand und während der Fahrt. Nach der Übung fuhren die Probanden eine Messfahrt je Messinstrument. Die Probanden wurden in zwei Gruppen geteilt. Die Hälfte der Probanden begann mit der DRT-Fahrt, die andere Hälfte mit der ICA-Fahrt. Die Streckenabschnitte wurden zudem in zwei verschiedenen Reihenfolgen aufgesetzt. Dabei begann die Hälfte der Probanden je Gruppe mit einer Fahrt durch die Abschnitte Landstraße - Stadt einfach - Stadt komplex. Die andere Hälfte durchfuhr in der ersten Messfahrt die Abschnitte in der Reihenfolge Stadt einfach - Stadt komplex - Landstraße. Die Abfolge der Nebenaufgaben sowie der Baselinefahrt je Abschnitt wurde randomisiert. Das Studiendesign war somit ein within-subjects design mit den unabhängigen Variablen Komplexität der Fahrszene sowie Art und Schwierigkeit der Nebenaufgabe. Als abhängige Variablen wurden die Reaktionszeit auf die DRT sowie der ICA gemessen. Außerdem wurde die Anzahl der bearbeiteten SuRT-Aufgaben sowie die Fehler in der n-back-task aufgezeichnet.

Ergebnisse

Zur Auswertung der ICA Daten wurden zunächst alle Werte einer Person über alle Bedingungen und Abschnitte z-standardisiert. Zur Überprüfung der Hypothesen wurden messwiederholte ANOVAs berechnet. Die Irrtumswahrscheinlichkeit wurde mit $\alpha = 5\%$ festgesetzt. Bei signifikanten Haupteffekten wurden post-hoc tests mit Bonferroni-Korrektur berechnet. Aufgrund von Extremwerten wurden die Daten eines Probanden aus der DRT-Auswertung ausgeschlossen. Bei der Aufzeichnung der ICA-Daten mussten die Daten eines Probanden aufgrund von technischen Problemen ausgeschlossen werden. Gemäß Hypothese 1 wurde untersucht, ob sich die Baseline-Bedingungen in den verschiedenen Streckenabschnitten unterscheiden. Die mittleren Reaktionszeiten sowie z-standardisierten ICA-Werte und Standardfehler sind in Tabelle 1 dargestellt.

Tab.1: mittlere DRT Reaktionszeiten und z-Werte des ICA mit Standardfehler in den Baseline-Bedingungen in ms.

	Baseline DRT	Baseline ICA	N
Landstraße	430 (15)	- 0,40 (0,06)	23
Stadt einfach	449 (17)	- 0,33 (0,04)	23
Stadt komplex	463 (13)	- 0,29 (0,04)	23

Eine ANOVA mit Messwiederholung zeigte einen signifikanten Haupteffekt der Fahrszene gemessen mit der DRT ($F(2,44) = 4.4$, $p < .05$, $\eta^2 = .17$). Die Reaktionszeiten in der DRT unterschieden sich signifikant zwischen einer nicht komplexen Fahrt auf der Landstraße ($M=430\text{ms}$, $SD= 67\text{ms}$) und einer komplexen Stadtfahrt ($M=463\text{ms}$, $SD= 60\text{ms}$). Der ICA konnte keinen signifikant unterschiedlichen Workload in Abhängigkeit der Fahrsituation messen ($F(2,44) = 2.09$, n.s.). Hypothese 1 kann daher nur für die DRT angenommen werden. Gemäß Hypothese 2 wurde untersucht, ob Bedingungen mit Nebenaufgabe mehr Workload erzeugen als Bedingungen ohne Nebenaufgabe. Außerdem wurde getestet, ob schwierige Nebenaufgaben mehr Workload erzeugen als einfache. Zunächst werden die Ergebnisse für die DRT Reaktionszeiten berichtet. Die mittleren Reaktionszeiten und Standardfehler sind in Tabelle 2 dargestellt

Tab.2: mittlere DRT Reaktionszeiten und Standardfehler der n-back und SuRT in ms.

	Baseline	0-back	2-back	SuRT einfach	SuRT schwer	N
Landstraße	430 (15)	501 (17)	681 (25)	610 (17)	648 (24)	23
Stadt einfach	449 (14)	537 (19)	674 (23)	627 (17)	644 (18)	23
Stadt komplex	463 (12)	576 (20)	665 (29)	672 (24)	676 (21)	23

Eine ANOVA mit Messwiederholung konnte in jedem Fahrtabschnitt einen signifikanten Haupteffekt der n-back-task nachweisen (Landstraße: $F(2,44) = 80.2$, $p < .001$, $\eta^2 = .79$; Stadt einfach: $F(2,44) = 69.3$, $p < .001$, $\eta^2 = .76$; Stadt komplex: $F(2,44) = 37.73$, $p < .001$, $\eta^2 = .63$). Beide Schwierigkeitsstufen erzeugen signifikant mehr Workload als die Baseline-Bedingung ($p < .001$). Sowohl auf der Landstraße ($p < .001$), als auch im einfachen Stadtabschnitt ($p < .001$) und im komplexen Stadtabschnitt ($p < .01$) erhöhten sich die Reaktionszeiten zudem signifikant mit zunehmender Schwierigkeit der n-back-task. Hypothese 2 kann damit für die DRT angenommen werden. Auch die DRT-Reaktionszeiten in der SuRT unterscheiden sich in jedem Abschnitt signifikant von der Baselinefahrt (Landstraße: $F(2,32) = 73.3$, $p < .001$, $\eta^2 = .77$; Stadt einfach: $F(2,44) = 101.9$, $p < .001$, $\eta^2 = .82$; Stadt komplex: $F(2,44) = 68.8$, $p < .001$, $\eta^2 = .76$). Bezüglich der SuRT konnte kein signifikanter Anstieg der Reaktionszeit mit zunehmender Schwierigkeit der SuRT ermittelt werden. Die mittleren Reaktionszeiten und Standardfehler der SuRT sind ebenfalls in Tabelle 2 dargestellt. Im Folgenden werden die Ergebnisse der ICA-Werte berichtet. Die mittleren z-Werte und Standardfehler für die Nebenaufgaben-Bedingungen sind in Tabelle 3 dargestellt.

Tab.3: z-standardisierte ICA-Werte und Standardfehler der n-back und des SuRT.

	Baseline	0-back	2-back	SuRT einfach	SuRT schwer	N
Landstraße	-.40 (.06)	-.36 (0,05)	-.29 (.05)	.34 (.10)	.29 (.06)	23
Stadt einfach	-.33 (.04)	-.25 (0,04)	-.20 (.08)	.65 (.07)	.37 (.05)	23
Stadt komplex	-.28 (.04)	-.21 (0,03)	-.21 (.07)	.57 (.06)	.40 (.06)	23

In Bezug auf die n-back-task konnte eine ANOVA mit Messwiederholung in keinem Fahrtabschnitt einen signifikanten Haupteffekt nachweisen (Landstraße: $F(2,44) = 2.4$, n.s.; Stadt einfach: $F(2,30) = 1.4$, n.s.; Stadt komplex: $F(2,44) = .67$, n.s.). Die durch den ICA indizierte Beanspruchung während der n-back-task unterscheidet sich daher weder gegenüber der Baselinefahrt noch zwischen den beiden Schwierigkeitsstufen der n-back-task. Es konnte lediglich ein tendenzieller Anstieg des Workload gezeigt werden. Hypothese 2 muss daher in Bezug auf den ICA abgelehnt werden. Hinsichtlich des SuRT konnte in allen Abschnitten ein signifikanter Haupteffekt gefunden werden (Landstraße: $F(2,44) = 28.5$, $p < .001$, η^2

= .56; Stadt einfach: $F(2,44) = 96.2$, $p < .001$, $\eta^2 = .81$; Stadt komplex: $F(2,44) = 71.2$, $p < .001$, $\eta^2 = .76$). Die SuRT-Bedingungen erzeugten in allen Abschnitten einen höheren Workload als die Baseline-Bedingungen ($p < .001$). In Bezug auf die beiden Schwierigkeitsstufen der SuRT zeigte sich, dass die leichte Variante in den beiden Abschnitten Stadt einfach und Stadt komplex einen höheren Workload erzeugte als die schwere ($p < .001$). Bei Betrachtung der Anzahl an bearbeiteten Aufgaben wird deutlich, dass diese in allen Fahrabschnitten in der leichten Bedingung signifikant höher ist als in der schweren ($F(1,21) = 86$, $p < .001$). In der leichten Variante wurden durchschnittlich 61 Aufgaben bearbeitet ($SD=19$), in der schweren lediglich 19 ($SD=9$). Wie aus Tabelle 3 hervorgeht, liegen die ICA-Werte in den SuRT-Bedingungen deutlich über denen der n-back-Bedingungen.

Diskussion

In dieser Studie sollten zwei Methoden zur Messung von kognitivem Workload miteinander verglichen werden, die Detection Response Task und der Index of Cognitive Activity. Beide Methoden wurden bisher in eher einfachen Fahrsituationen verwendet. Es sollte untersucht werden, ob beide Methoden auch in komplexeren Fahrszenarien anwendbar sind, da diese die tatsächliche Fahraufgabe besser abbilden. Dafür wurde eine Strecke mit unterschiedlich komplexen Abschnitten erstellt, in denen Nebenaufgaben mit ansteigender Komplexität bearbeitet werden mussten. Es konnte gezeigt werden, dass die DRT ein robustes Instrument zur Messung von kognitivem Workload darstellt, welches auch in komplexen Fahrsituationen Beanspruchung zuverlässig messen kann. In der kognitiven n-back-task zeigten sich in allen Fahrabschnitten signifikante Unterschiede in Abhängigkeit der Fahrsituation und Komplexität der Aufgabe. Auch die Fahrt mit SuRT-Bedienung erzeugte mehr Workload als eine Fahrt ohne Nebenaufgabe. Allerdings konnte die DRT weniger deutlich Unterschiede hinsichtlich der Komplexität des SuRT messen. Dieses Ergebnis ist erwartungskonform, da die SuRT visuell-manuelle Beanspruchung erzeugt und die DRT vorgibt, kognitiven Workload zu messen (Engström, Johansson & Östlund, 2005). Der ICA konnte weder einen signifikanten Unterschied in Abhängigkeit der Komplexität der Fahrszene noch der Komplexität der kognitiven Nebenaufgabe messen. Es ergaben sich zudem weitere unerwartete Effekte. Zum einen erzeugte die leichte Variante der SuRT einen signifikant höheren ICA als die schwere Variante. Dies könnte daran liegen, dass die Probanden in der leichten SuRT Variante deutlich mehr Aufgaben bearbeiteten, das heißt selbstständig ihre Arbeitslast erhöhten. Dies wurde möglicherweise durch den hohen Aufforderungscharakter der einfachen SuRT hervorgerufen, so viele Aufgaben wie möglich erfolgreich bearbeiten zu wollen. Es kann daher vermutet werden, dass der ICA nicht dazu geeignet ist, in einem Setting mit komplexen Fahrszenarien Aufgaben mit unterschiedlichem Schwierigkeitsgrad zu diskriminieren. Vielmehr wird mit dem ICA eher generelle Arbeitslast gemessen, die nicht nur durch die Komplexität einer Aufgabe bestimmt wird. Zum anderen lagen die ICA-

Werte für die Bearbeitung der visuell-manuellen Nebenaufgabe deutlich über denen der kognitiven Nebenaufgabe, was ebenfalls nicht den Erwartungen entspricht, da der ICA genau wie die DRT vorgibt, kognitive Ablenkung zu messen (Marshall, Davis & Knust, 2004). Eine mögliche Ursache für die hohen SuRT-Werte könnte auch darin liegen, dass bei der Bearbeitung des SuRT viele Blickwechsel von der Straße weg hin zum Display im Fahrzeug nötig sind. Der ICA-Wert könnte durch diese Blickwechsel beeinflusst werden. Es scheinen daher weitere Validierungsversuche nötig zu sein, um mögliche Artefakte des ICA in einem anwendungsnahen Setting ausschließen zu können. Ein möglicher folgender Ansatz könnte darin bestehen, zum Einen die Anzahl der zu bearbeitenden SuRT-Aufgaben und zum Anderen die Lage des SuRT-Screens konstant zu halten, um diese Artefakte untersuchen zu können.

Literatur

- Alliance of Automobile Manufacturers (2006). Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems, Washington, D.C.: Alliance of Automobile Manufacturers.
- Conti, A., Dlugosch, C., Vilimek, R., Keinath, A. & Bengler, K. (2012). An assessment of cognitive workload using detection response tasks. In Stanton, N.A. (Ed.), *Advances in human aspects of road and rail transport* (pp. 735-743). Boca Raton, FL: Taylor & Francis Group.
- Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F*, 8 (2), 97-120.
- Fastenmeier, W. (1995a). Die Verkehrssituation als Analyseeinheit im Verkehrssystem. In W. Fastenmeier (Eds.), *Autofahrer und Verkehrssituationen: Neue Wege zur Bewertung von Sicherheit und Zuverlässigkeit moderner Straßenverkehrssysteme*. Köln: Verlag TÜV Rheinland.
- ISO/TS 14198:2012. Road vehicles - Ergonomic aspects of transport information and control systems - Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems
- Jahn, G., Oehme, A., Krems, J.F. & Gelau, C. (2005). Peripheral Detection as a Workload Measure in Driving: Effects of Traffic Complexity and Route Guidance System Use in a Driving Study. *Transportation Research Part F* 8, 255-275.
- Marshall, S.P., Davis, C. & Knust, S. (2004). The Index of Cognitive Activity: Estimating Cognitive Effort from Pupil Dilation. Technical Report ETI-0401. Eyetracking Inc., San Diego.
- Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. (2009). Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Per-

formance in Young Adult Drivers. *Transportation Research Record: Journal of the Transportation Research Board*, 2138, 6-12.

Rösler, D. (2010). Fahrrelevanz von Elementen des Straßenverkehrs – Beschreibung, Erfassung und Anwendung. (Dissertation). Chemnitz: Technische Universität Chemnitz.

Schwalm, M., Keinath, A. & Zimmer, H.D. (2008). Pupillometry as a Method for Measuring Mental Workload within a Simulated Driving Task. In: de Waard, D., Flemisch, F.O., Lorenz, B., Oberheid, H., Brookhuis, K.A. (eds.) *Human Factors for Assistance and Automation*, pp. 1–13. Shaker Publishing, Maastricht.

Schwalm, M. (2009). Pupillometrie als Methode zur Erfassung mentaler Beanspruchungen im automotiven Kontext. (Dissertation). Saarbrücken: Universität Saarbrücken.

Vilimek, R., Schäfer, J., & Keinath, A. (2013). Effects of Task and Presentation Modality in Detection Response Tasks. *HCI* (16) 2013: 177-185.

Strategies for Efficient Driving with Electric Vehicles

Magnus Helmbrecht¹, Klaus Bengler¹ and Roman Vilimek²

Keywords: Electric Vehicles, Driving Efficiency, Driver Behavior, HMI

Abstract

This paper describes the aspect of increasing driving efficiency with electric vehicles by optimizing driving behavior. A field study with MINI E electric vehicles was set up to investigate the saving potential of efficient driving strategies. In this study 27 experienced MINI E users were driving a 32 km reference route, once according to their normal driving behavior and a second time with the instruction to drive as efficiently as possible. Analysis of the driving dynamics data showed that participants generally applied valid strategies to significantly reduce energy consumption. Nevertheless, specific traffic situations which require assistance for optimal efficient driving could be identified.

Introduction

Electric Vehicles (EV) are seen as a critical contribution to emission-free future mobility. However, intelligent energy management is an essential aspect since the battery capacity is a trade-off between cost and everyday mobility needs. While technological optimization of drivetrain, power electronics and battery management advances, driving style will remain a major source of influence. A field study with the MINI E electric vehicle was set up to investigate efficient driving strategies for EVs in order to quantify the driver-dependent saving potential and the need for driver assistance by driver assistance systems (DAS) or in-vehicle information systems (IVIS).

The Driver's Influence on Driving Efficiency

According to Dorrer (2003), the fuel consumption of combustion vehicles is significantly dependent on driving behavior. Efficient driving strategies may reduce fuel consumption of internal combustion engine (ICE) vehicles by 26% to 58% depending on road type. Under the aspect of anticipatory driving, the following strategies have proved to be very successful (Popiv, Rommerskirchen, Bengler, Duschl and Rakic, 2010):

- Avoiding unnecessary acceleration and unnecessary high final velocity
- Long coasting, i.e. decelerating with the vehicle's gained momentum without propulsion, to avoid braking

¹ Technische Universität München, Institute of Ergonomics, Garching, Germany

² BMW Group, Concept Quality, Munich, Germany

- Furthermore the powertrain of ICEs requires a certain load for efficient running (Fiala, 2006), so that another aspect is:
- Fast and powerful acceleration to reach the vehicle's optimum operating point as fast as possible

While driving strategies for ICEs have almost become general knowledge in the area of efficiency improvement, this is different in regards to EVs. Although some research has been done to identify driving-style-dependent energy saving potential (e.g. Walsh, Carroll and Eastlake, 2011), few general statements on efficient driving strategies for EVs applicable in real traffic situations have been made until now. From a technical perspective, efficient driving with EVs is quite similar to that with ICEs. The aspect of anticipatory driving is based on a physical principle and is therefore independent of the type of powertrain. Also, coasting as much as possible is relevant for EV drivers since regenerative braking only works with a technically limited efficiency factor; so every regenerative braking maneuver is accompanied by loss. The main difference in efficient driving with EVs concerns the acceleration behavior. Combustion engines offer a very small efficiency factor at lower speed and lower load, therefore a strong acceleration to cruising speed is very efficient. In contrast, the EV engine operates optimally in a higher velocity and load spectrum (AC Propulsion, 2013). Hence, in common application cases smooth accelerating is more efficient than the powerful acceleration strategies known from combustion vehicle driving. In summary, the most successful strategies for efficient driving with electric vehicles are the following:

- Avoiding unnecessary acceleration and unnecessary high final velocity
- Long coasting to avoid braking
- Slow and smooth acceleration

Driver Assistance and Information Systems for Efficient Driving

In general DAS and IVIS have to fulfill basic requirements for safety, usability and user acceptance. Usability and acceptance depend on the constellation of driving situation and driver's intention and are a measure for the system's success. The main aspect of safety is that the system should not distract the driver's attention from the road scene.

The principal goals of today's assistance systems in ICEs are the improvement of safety, the increase of comfort and the optimization of driving efficiency. According to Popiv, Rakic, Bengler and Bubb (2009), supporting the driver in anticipatory driving is a very promising approach. It has been proven that unnecessary braking and accelerating can be avoided by extending the driver's natural anticipation horizon (Rommerskirchen, Helmbrecht and Bengler, 2013).

Several IVIS are already implemented in today's EVs. For example, the Nissan Leaf displays the current range radius on the onboard navigation system and provides an overview of the energy used by electric motor, climate control and other

systems (Nissan, 2013). The BMW ActiveE also provides advanced energy monitoring functions and allows an increase in energy efficiency by switching on the ECO PRO driving mode. The BMW i3 will additionally feature a spidergram that can be superimposed on the navigation system map. It indicates which destinations the user can reach with the current battery charge level and provides an estimate of additional range potential achievable with efficient driving. These systems provide specific feedback on the driver's economy, but a huge additional improvement potential lies in actively assisting the driver in performing efficiency optimized driving strategies.

Research Questions

Increasing driving efficiency of EVs by specific driving behavior is a rather new topic of research in electric mobility. Up until now there has not been much knowledge about the influence of specific driving strategies on the energy consumption of EVs. Therefore the basic question is: what is the quantitative energy saving potential that can be reached by an economical driving style? Since EV power consumption depends on required acceleration and driving speed, the driver should be able to reduce the vehicle's energy consumption significantly by a purposeful, efficient driving style. In this context, the second question is: which driving strategies do the participants implement to drive economically? Since all subjects of this study are long-term EV users, they would probably have enough driving experience to be capable of applying correct EV-typical saving strategies that differ somewhat from those for ICE. However, it will have to be determined whether the theoretically available savings potential is completely utilized. On the one hand, the environmental factors of traffic situations or the limited human anticipation horizon might hinder the driver from driving perfectly efficient. On the other hand, misunderstandings concerning EV power characteristics could aggravate the application of efficient driving strategies. This leads to the questions: how can the driver be assisted to drive the EV efficiently and how could IVIS or DAS assist the driver to fulfill this requirement?

Methods

The EV used for this field test was the MINI E, a two-seater conversion electric vehicle based on the familiar MINI hatch. The MINI E was specifically designed to conduct studies focusing on customer experience and customers' requirements as part of the BMW i development roadmap. Details on the MINI E research projects can be found in Vilimek, Keinath and Schwalm (2012). The integration of field trial results in the BMW customer-centered design process is described in Vilimek and Keinath (accepted for publication).

The installed e-machine delivers a near constant torque of 220 Nm and a peak power of 150 kW. This enables the driver to engage in a wide spectrum of possible driving behavior - from very sporty to very efficient. Positive accelerations in

common driving are possible up to 0.4 g, regenerative braking by releasing the accelerator pedal up to 0.24 g. The velocity is electronically limited at 152 km/h. The reference energy consumption is indicated as 15 kWh/100 km. In real terms, and depending on the driving style, the MINI E's range is about 160 km. Charging with a 32 ampere wall box takes about 3.8 hours. Using a conventional plug (12 ampere, 240 volts), about 10.1 hours are needed for a full charge from 0% to 100%. For the purpose of this study, the test vehicles were equipped with instruments for measuring driving dynamics.

27 persons took part in this field study. All of them had the MINI E in use for at least two months at the time of test; 15 of them for private purposes, 12 drove it as part of their company car pool. The mean age was 43.4 years ($SD = 9.8$), three of the customers were female (11%). All subjects had MINI E driving experience of more than 500 km. Post-analysis of recorded data revealed no relevant differences in driving experience.

A 31 km course north of Munich was selected as the test route. Passing through towns and continuing along country roads and highways, this track contains a variety of common driving situations, according to Fastenmeier (1995). These range from simple road intersections or short highway trips to multi-lane traffic light junctions. Among other things, the numerous accelerating and decelerating operations for speed limits and traffic lights are particularly important for analyzing driver specific energy saving potential. In order to enable a road-type-specific data analysis, the test route was divided into 5 sections from inner-city to highway (Fig. 1).

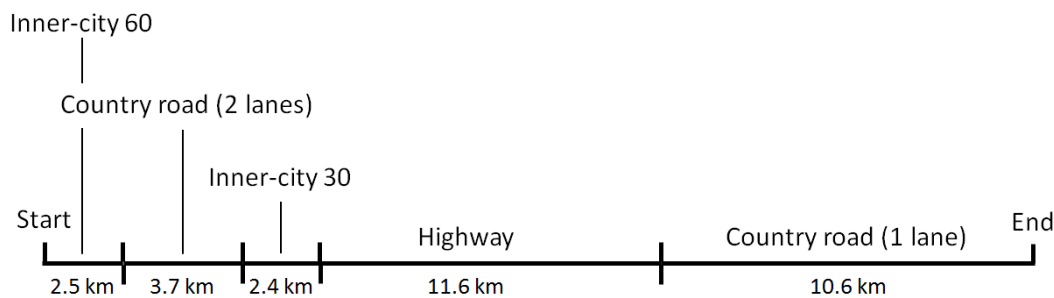


Fig. 1: Sections of the Test Route

The test route was driven two times by each of the private and corporate MINI E users. During one of these two trips, the participants received no special instruction. Concerning the other trip, the subjects were told to drive as efficiently as possible in order to reach the end of the route with a minimum of energy consumption. The participants were accompanied by observers in the passenger seat on both drives. Therefore a detailed monitoring of the in-test traffic situations and specific driver reactions (e.g. pedal activity) was possible. In order to avoid learning effects during the running test drives, the order of the two drives was permuted with the subjects. Each drive was started with a fully charged battery. This ensured a comparable initial status of the test vehicle and made it possible to calculate the con-

sumed electric energy by determining the battery's energy content after the test drive.

Driving dynamics data was collected with a smartphone (Apple iPhone 4) in test vehicle during both drives. Using the smartphone's sensor technology, acceleration was recorded at a frequency of 25 Hz, position and velocity out of the GPS signal at a frequency of 1 Hz. The smartphones were positioned in the test vehicles with one of their coordinate axes exactly parallel to the longitudinal axis of the vehicle. An appropriate fastening mechanism prevented uncontrolled movement of the measuring devices.

Road-type dependent driving strategies have become recognizable by creating location-related speed and acceleration profiles from the smartphones' raw data. Therefore driving profiles were generated for each participant in each section of the test route. An individual comparison of equivalent speed and acceleration profiles showed differences in driver behavior as a result of different driving strategies applied in the respective section. In order to verify differences in the profiles by applying statistical methods, efficiency-related values were defined and calculated for each profile in each section. These are: average speed, standard deviation of speed, maximum speed, average positive acceleration, average negative acceleration (deceleration), and standard deviation of acceleration.

Results

A t-test for dependent groups was used for the statistical analysis of differences between normal and efficient driving. Standard significant levels were chosen (5%). Since the batteries' energy content cannot be reliably determined while the car is driving, a value of consumed energy only exists for the complete test route, but not for each section. By applying strategies for efficient driving, the participants significantly reduced the average energy consumption from 17.41 kWh/100km (SD = 2.42) to 12.74 kWh/100km (SD = 1.82), $t(26) = 8.527$, $p < 0.001$. This proves a saving potential by driving behavior of 26.8% for electric vehicles in a realistic scenario in practice. In comparison with empiric values according to Dorrer (2003), the fuel savings potential for combustion vehicles for an assimilable route is estimated to be slightly higher at about 35%. Although there are no exactly comparable values for internal combustion vehicles on this route, the influence of different power characteristics and regenerative braking is noticeable in this context. Nevertheless, even with current EVs without assistance, a significant reduction of consumed energy is possible by strategic driving. Similar to efficient driving with internal combustion vehicles, this does not equate with longer trip duration. The mean traveling time in the efficient driven lap ($M = 38.5$ min, $SD = 2.7$) was only one minute higher than in normal cruise ($M = 37.2$ min, $SD = 8.3$), which is not significant, $t(26) = -0.852$, $p = 0.402$.

The detailed analysis of efficiency-characteristic values shows that the applied strategies for efficient driving vary with road type (section of test route). Although

a general endeavor to accelerate and decelerate smoother becomes obvious by a significantly smaller standard deviation of acceleration for the efficient lap ($M = 0.065 \text{ m/s}^2$, $SD = 0.010$) than for the normal lap ($M = 0.048 \text{ m/s}^2$, $SD = 0.007$) on the whole test route, $t(26) = 9.292$, $p < 0.001$. The average speed was not significantly lower in every section. In inner-city-30, where the average speed of the normal lap ($M = 27.31 \text{ km/h}$, $SD = 3.12$) did not differ significantly from that of the efficient lap ($M = 25.58 \text{ km/h}$, $SD = 3.84$), $t(26) = 1.947$, $p = 0.062$. This does not come as a surprise, since a speed limit of 30 km/h does not leave much room to drive more slowly without being a hindrance to traffic. This assumption is confirmed by the subjective assessment of test subjects. The faster sections inner-city-60 and country road (2 lanes) are different cases. In principle, both offer the ability of efficient anticipatory driving. However, the large number of traffic lights in these situations seems to hinder the drivers from efficient anticipatory driving. This is reflected in very small mean average speed differences between normal lap in country road (2 lanes) ($M = 52.19 \text{ km/h}$, $SD = 8.63$) and efficient lap in country road (2 lanes) ($M = 49.45 \text{ km/h}$, $SD = 9.28$). These differences are not significant, $t(26) = 1.180$, $p = 0.249$. In fact, the average positive acceleration and average deceleration in the normal lap were already extraordinarily high in these sections. In the efficient lap these values were smaller, but still on the same level as those of the normal lap in other sections of the test route. In contrast to that finding, the average speeds in the section highway were significantly smaller in the efficient driven lap ($M = 76.29 \text{ km/h}$, $SD = 6.80$) than in the normal lap ($M = 76.29 \text{ km/h}$, $SD = 8.55$), $t(25) = 8.450$, $p < 0.001$. This is the same in the second exurban section without traffic lights country road (1 lane), where the average speed of the normal lap ($M = 61.13 \text{ km/h}$, $SD = 5.42$) was also significantly higher than that of the efficient lap ($M = 54.55 \text{ km/h}$, $SD = 6.06$), $t(26) = 4.511$, $p < 0.001$.

To sum up, the participants of this field trial generally applied valid strategies for efficient driving with electric vehicles. They accelerated and decelerated in a smooth manner and reduced their average speed on highways and country roads. Nevertheless, situations with traffic lights aggravated the application of efficient driving strategies.

Opportunities to Support Efficient Driving with Electric Vehicles by DAS and IVIS

In principle drivers of EVs could be assisted in anticipatory driving by DAS to gain the maximum of saving potential by enlarging the human anticipation horizon. Whether this makes sense or not - especially under the aspects of acceptance and safety - has to be investigated in further studies. According to subjective assessment during the test drives, IVIS dynamically informing the driver when and on which level efficient driving strategies are required to reach the destination would be a useful measure to improve range in regards to electric vehicles. However, the most promising application area for assistance systems lies in optimizing

driving efficiency by providing traffic light assistance, especially for country roads. The following example (Fig. 2) illustrates the potential of anticipatory traffic light assistance. In this section traffic lights are at km 0.1, km 0.8, km 1.4, and at km 2.2. The profiles of 'efficient drive' and 'normal drive' are generated from the measured values of a representative subject. The profile 'optimal with DAS' is the theoretical optimum that can be reached by driver assistance.

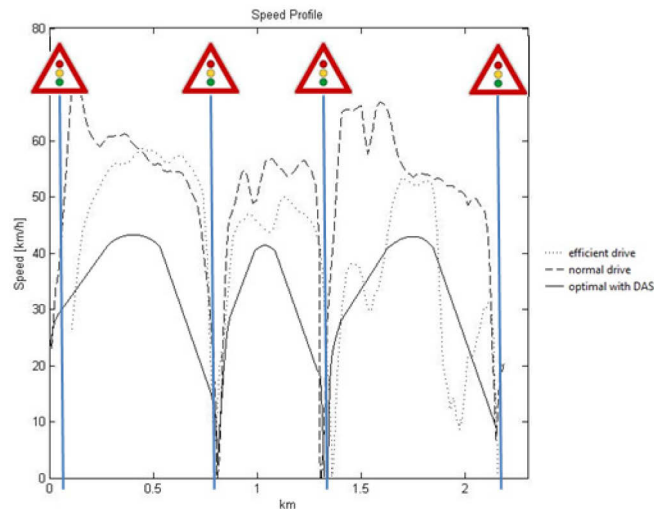


Fig. 2: Saving potential of traffic light assistance

Instead of strong acceleration after the first traffic light turned green and then decelerating hard in front of the next red light, an optimal acceleration and optimal final speed would avoid unnecessary energy transmission and loss. Krause and Bengler (2012) have analyzed the saving potential of such traffic light assistances in ICE and have reported 20% less fuel consumption. Equivalent test drives in real traffic must be done for EVs in future studies.

Conclusion

The comparison of normal and efficient driving behavior on the test route showed that participants were able to reduce their EV's energy consumption by more than 25%. This means that at least two months and 500 km of driving experience are sufficient to apply efficient EV-typical driving strategies. However, in some traffic situations there is still untapped potential for energy saving. As well as the drivers of ICEs, the EV drivers of this study require assistance for anticipatory driving. To enable this by DAS or IVIS should be the goal of further research and development in electric mobility.

Acknowledgements

The authors would like to acknowledge the Bavarian Ministry of Economic Affairs, Infrastructure, Transport and Technology and the BMW Group for making this research possible. Special thanks to the team members of project i at BMW

and eMUC at TUM for their valuable contributions and to Julia Messerschmidt, Christina Schöberl and Karin Thomsen for helpful comments on an earlier draft of this document.

Literature

- AC Propulsion (2013). Datasheet: AC-150 Gen-2 EV Power System. url: <http://www.acpropulsion.com/datasheet/ac150gen2.pdf>
- Dorrer, C. (2003). Effizienzbestimmung von Fahrweisen und Fahrerassistenz zur Reduzierung des Kraftstoffverbrauchs unter Nutzung telematischer Informationen. Stuttgart: Dissertation Universität Stuttgart.
- Fastenmeier, W. (1995). Autofahrer und Verkehrssituationen - Neue Wege zur Bewertung von Sicherheit und Zuverlässigkeit moderner Straßenverkehrssysteme. Köln: Verlag TÜV Rheinland.
- Fiala, E. (2006). *Mensch und Fahrzeug*. Wiesbaden: Vieweg+Teubner Verlag.
- Krause, M. & Bengler, K. (2012). Traffic Light Assistant - Driven in a Simulator. *Proceedings of the 2012 IEEE Intelligent Vehicles Symposium Workshops*.
- Nissan (2013). url: <http://www.nissanusa.com/electric-cars/leaf/charging-range/range/>
- Popiv, D., Rakic, M., Bengler, K. & Bubb, H. (2009). Timing Concept for Assistance of Anticipatory Driving. *Proceedings of the 17th World Congress on Ergonomics IEA*.
- Popiv, D., Rommerskirchen, C., Bengler, K., Duschl, M. & Rakic, M. (2010). Effects of assistance of anticipatory driving on driver's behaviour during deceleration situations. *Proceedings of the 2nd European Conference on Human Centred Design of Intelligent Transport Systems (HUMANIST)*.
- Rommerskirchen, C., Helmbrecht, M. & Bengler, K. (2013). Increasing complexity of driving situations and its impact on an ADAS for anticipatory assistance for the reduction of fuel consumption. *Proceedings of the 2013 IEEE Intelligent Vehicles Symposium*.
- Vilimek, R., Keinath, A. & Schwalm, M. (2012). The MINI E field study - Similarities and differences in international everyday EV driving. In Stanton, N.A. (Ed.), *Advances in human aspects of road and rail transport* (pp. 363-372). Boca Raton, FL: CRC Press.
- Vilimek, R. & Keinath, A. (accepted for publication). User-centred design and evaluation as a prerequisite for the success of disruptive innovations: An electric vehicle case study. In M. Regan, T. Horberry & Stevens, A. *Driver acceptance of new technology: Theory, measurement and optimization*. Aldershot: Ashgate.
- Walsh, C., Carroll, S. & Eastlake, A. (2011). UK Electric Vehicle Range Testing and Efficiency Maps. SAE International, Technical Papers. url: <http://papers.sae.org/2011-39-7224/>

Effects of telemedical bystander support on helping behaviour after traffic accidents

Katja Karrer-Gauß¹ and Matthias Issing²

Keywords: telemedical system, helping behaviour, bystander behaviour

Abstract

Witnesses of traffic accidents often do not use the time before arrival of the emergency team to help accident victims. This failure to show helping behaviour may be rooted in fears or insecurities. Some of the environmental factors that inhibit helping behaviour can be resolved through the use of a telemedical system. In a survey of 132 persons, it was shown that in some cases the provision of a telemedical emergency kit increased spontaneous helpfulness. There was a marked increase among those individuals whose subjective or objective first aid ability was minimal and who feared being infected or making mistakes in the absence of telemedical support. Our results suggest that the use of telemedical systems could significantly increase the success rate in saving lives in traffic accidents.

1 Introduction

In 2011, 4,009 people died as a result of traffic accidents in Germany and 68,985 were severely injured (Statistisches Bundesamt 2012). For accident victims with life-threatening injuries, the span of time from the occurrence of the accident to the provision of first aid (the “therapy-free interval”) is crucial in determining the chances of survival and the success of later treatment (Schwermann et al. 2003).

Germany has a well-developed ground and air-based first aid system: it takes an average of only 13 minutes in cities and 21 minutes in rural areas for emergency teams to arrive (Central Union of Electrotechnology and Electrotechnical Industry e.V. 1999). In Germany, this time frame has been regulated by law: As a guiding value, the emergency team is expected to arrive in 95 percent of all cases within 8 to 17 minutes, varying from state to state (Forplan Dr. Schmiedel, 2013).

In the case of critical emergencies, however, this is still too long. After cardiovascular arrest, the probability of survival without effective first aid drops at a rate of 10% per minute (Schlechtriemen et al. 2003).

The therapy-free interval can be shortened when bystanders at the scene of the accident (usually medical laypersons) provide effective first aid. In the case of critical emergencies, the conditions for helping accident victims are much better for bystanders than for the emergency team due to the time advantage. To help victims until the emergency team arrives, it is crucial on the one hand that bystanders take

¹ Technische Universität Berlin, Human-Machine-Systems, MAR 3-1, Marchstr. 23, D-10587 Berlin, Germany

² medneo GmbH, , Marie-Elisabeth-Lüders-Str. 1, D-10625 Berlin, Germany

the initiative and are willing to help, and on the other, that they are capable of actually providing help of relatively high quality. Studies show that many potential helpers are often present at accidents, but that they seldom actually provide help (Donner-Banzhoff et al. 1999).

The success rate for saving lives could be significantly improved by increasing the willingness of bystanders to provide first aid. One potential means of doing so is by making telemedical technologies available that can assist laypersons in administering first aid at the site of the accident. Telemedicine is the use of telecommunication and information technologies in order to provide clinical health care at a distance.

1.1 Helpfulness, helping behavior and factors inhibiting helpfulness

Bystanders' willingness to help is the prerequisite for providing first aid to victims of traffic accidents. It is mainly in the discipline of social psychology that helpfulness has been investigated and possible explanations for helping behaviour explored.

A well-known model of helping behaviour is the decision model by Latané and Darley (1969, 1970), which describes five different stages a potential helper goes through before actually engaging in helping behaviour. In deciding whether to help, the bystander has to be aware of the accident (phase 1), interpret it as an emergency situation (phase 2), realize that it is his or her responsibility to help (phase 3), decide what kind of help he or she can offer (phase 4), and then decide whether to help or not. The model reveals a number of reasons why people decide not to help. In each phase, the decision process can be terminated with no resulting helping behaviour. Piliavin et al. (1981) argue that helping behaviour depends on the costs and benefits associated with the act of helping or not helping: the costs of helping may include the effort, possible negative experiences, and physical harm, and the benefits may include praise from others. The costs of not helping may include self-criticism and being criticized by others, and one benefit may be that one does not need to disrupt one's own activities.

Walz (1991) suggested grouping the inhibiting factors into three categories: characteristics of the witness, characteristics of the situation, and characteristics of the victim. One of the characteristics of the witness that inhibits the willingness to help is a deficit in self-perceived competence. Relations between competence and willingness to help have been identified repeatedly (Form & Nosow 1958, Midlarsky 1971, Midlarsky & Midlarsky 1973, Schwartz & David 1976, Kazdin & Byran 1971, Ashton & Severy 1976, Huston et al. 1981, Rayko 1977). One characteristic of the situation inhibiting helping behaviour is the ambiguity of the situation, i.e., uncertainty as to whether it actually is an emergency (Latané & Darley 1969, Clark & Word 1972, Gaertner & Dovidio 1977, Shotland & Heinold 1985). Another well-researched situational factor is the presence of other bystanders (Shotland & Heinold 1985). The individual willingness to help decreases with the

size of the bystander group due to the diffusion of responsibility. Characteristics of the victim that inhibit the willingness to help may, for example, involve processes of causal attribution. Here, if a victim is seen as innocent he is more likely to receive help than if he himself caused the accident (Batson 1998). Unpleasant characteristics of the victim such as the presence of blood may also inhibit helping behaviour due to fear of infectious diseases, etc. (Brenner et al. 1997, Locke et al. 1995, Brenner & Kauffman 1993, Ornato et al. 1990).

1.2 Ways of enhancing helping behaviour

There are several ways of enhancing bystanders' spontaneous willingness to help by promoting the factors that facilitate helpfulness and reducing the ones that inhibit it. These include training laypersons in first aid methods, conveying information about emergency situations, and offering telemedical instruction to accident witnesses.

It has been shown repeatedly that people behave more carefully and, in an emergency, help faster and more effectively after first aid training (Larsson et al. 2002). The general probability to offer help is also significantly higher for people who have taken part in a training program (Peterson & Russel 1999). In several experimental studies, it has been found that conveying information about emergency situations (such as traffic accidents) in a training context further enhances helpfulness (Beaman, Barnes, Klentz & McQuirk 1978, Pantin & Carver 1982, Rayko, 1977). Pantin & Carver (1982) found that showing videos of first aid instruction reduced people's reaction time before offering help.

Another possible means of enhancing helpfulness after traffic accidents is through telemedical instruction to accident witnesses in administering first aid. Through mobile audio and video communication with a medical call centre, laypersons can potentially be guided by experienced experts in providing first aid. Here, the parallel transmission of data obtained through non-invasive biomedical sensors enables call centre experts to get a detailed picture of the medical condition of the injured person at the site of the accident.

The easiest method of providing telemedical first aid instruction to laypersons is by telephone. Instruction by telephone in cardiovascular emergencies is already being used in many emergency units in the United States and Europe. Positive effects of telephone instruction have been found in several studies (Carter et al. 1984, Kellermann et al. 1989). The willingness of laypersons to begin administering cardiopulmonary resuscitation (CPR) before the arrival of the emergency team was enhanced in particular through telephone instruction, raising the rate of helpfulness to 54% from a former 32% (Culley et al. 1991). A large-scale survey indicated that 76% of accident witnesses who call an ambulance expect telephone support in how to behave until the emergency team arrives at the site of the accident (Billittier et al. 1999). However, even with instruction over the phone, factors like emotional

arousal, fear of infection, unpleasant characteristics of the victim, and medical or legal considerations still have an inhibiting influence (Hauff et al. 2003).

1.3 Aim of the study

The present study aims at analysing the factors that influence the helpfulness of laypersons coincidentally present at traffic accidents. The emphasis is on factors that can be influenced through use of a telemedical system. Our estimation of the costs and benefits of helping behaviour suggests that a telemedical system can contribute to reducing the costs because it eliminates or decreases the fear associated with the helping behaviour (reduction of fear of mistakes, fear of infection etc.). To investigate this assumption, we analyse whether the availability of a telemedical system can enhance helpfulness. Moreover, we identify the specific factors responsible for a change in behaviour and their respective importance. From the strength of salient factors, we will formulate suggestions for the design of telemedical systems currently under construction.

2 Method

For the survey, a fully standardized questionnaire was used describing two different traffic accident scenarios involving severely injured persons.

In the scenarios, the participants were asked to imagine being present at the scene of the accident before arrival of the emergency team. In the first scenario, the only thing available to the bystanders was the first aid kit that every vehicle has to be equipped with by law. In the second scenario, the participants were asked to imagine that they also had a telemedical first aid kit at their disposal. This is an expanded first aid kit with technical equipment allowing the witness of the accident to establish a direct connection (through a video and audio transfer system) to a doctor in a call centre who supervises the helping process.

After listening to the presentation of the two scenarios, the participants were asked to estimate their spontaneous helpfulness and to indicate the extent to which they agreed or disagreed with a series of statements on a six-point rating scale. The answers were used to assess each participant's parameter value regarding the factors that influence helping behavior and were chosen on the basis of a literature review. Factors that might influence willingness to help after traffic accidents in the scenario without telemedical equipment were helpfulness, fear of mistakes, fear of legal consequences, fear of putting oneself at risk, fear of infection, feeling overstrained, social obligation to offer first aid. In the scenario with telemedical equipment factors that influence willingness to help after traffic accidents were enhanced helpfulness, no longer feeling overstrained, no more fear of mistakes, no more fear of legal consequences, reduced feeling of putting oneself at risk, reduced fear of infection, openness to guidance, reduced fear of failure, attitude towards electronic equipment.

The following demographic characteristics were assessed in the questionnaire: age and gender, training in first aid methods (medical profession or education, simple or advanced first aid training, time elapsed since training), rating of respondent's first aid ability (general first aid ability and, more specifically, cardiopulmonary resuscitation) and experience with emergency situations (having witnessed or experienced emergencies, having witnessed or experienced traffic accidents),

To analyse the data, a bivariate analysis was performed. For the bivariate analysis, the Pearson correlation coefficients were calculated between the influencing factors of interest and the variable helpfulness (i.e., also enhancement of helpfulness).

3 Results

3.1 Participants

132 participants ranged in age from 17 to 83 years (47.7% women and 52.3% men). Of these participants, 92.4% had taken part in first aid training at least once. Slightly more than half had taken a course in lifesaving methods, and 44.3% had gone through extensive first aid training. Altogether, 13.6% had been trained in a medical profession. Despite the very high percentage of people with first aid training, only 5.3% of the participants said they exactly knew how to perform cardiopulmonary resuscitation. Furthermore, 17.4% said they did not know what to do at all. 39.4% of the participants had experienced at least one emergency situation in which a person needed first aid for a severe injury or medical emergency. 65.4% of these situations were the consequence of a traffic accident.

3.2 Factors influencing helpfulness

The individual ratings of specific influencing factors were rated on a scale from 0 (strongly disagree) to 5 (strongly agree) for the emergency scenario without the telemedical kit.

Mean values of the factors indicate that feeling overstrained (1.6), fear of legal consequences (1.8), and fear of putting oneself at risk (2.0), as well as fear of infection (1.8) tended not to be seen as very important, whereas the statement first aid as a social obligation with a mean value of 4.7 was clearly important. About 30% of the participants rated their helpfulness in the scenario as high, whereas only 1.5% rated their helpfulness as low.

In a bivariate analysis, the most important factor possibly preventing helpfulness was fear of mistakes, manifested in a highly significant and negative correlation ($r = -0.42$, $p < .001$). A highly significant and positive correlation was found between helpfulness and subjective self-rating of cardiopulmonary resuscitation ability ($r = 0.40$, $p < .001$) and general first aid ability ($r = 0.37$, $p < .001$). With a positive atti-

tude towards social obligation to help, helpfulness was also rated higher ($r = 0.35$, $p < .001$). Other inhibiting factors showed highly significant and negative correlations with helpfulness: feeling overstrained ($r = -0.33$, $p < .001$), fear of legal consequences ($r = -0.33$, $p < .001$), fear of infection ($r = -0.32$, $p < .001$) and fear of putting oneself at risk ($r = -0.32$, $p < .001$). No significant correlation was found concerning the variables age, medical profession or education, time elapsed since most recent first aid training, and having witnessed or experienced emergencies.

3.3 Influencing factors on enhancing helpfulness

In the questionnaire, it was explained how the telemedical system works. Respondents were asked whether the availability of such a system would lead to an enhancement of helpfulness. Then the influencing factors that could possibly be responsible for enhancement of helpfulness were assessed again.

With a mean value of 3.85, most participants agreed that their helpfulness would be enhanced if a telemedical kit were available. Participants also supported the statements of no longer feeling overstrained by first aid methods ($M = 3.46$), and of being open to telemedical guidance ($M = 4.53$). The majority of participants did not agree that telemedical guidance could lead to fear of not being able to follow the instructions correctly and therefore failing in front of others ($M = 1.27$).

In the bivariate analysis of the influencing factors on helpfulness given the availability of a telemedical kit the factor openness to guidance showed the strongest and highly significant correlation with enhancement of helpfulness ($r = 0.61$, $p < .001$). Also, almost all of the factors showed highly significant and positive correlations with the enhancement: no longer feeling overstrained ($r = 0.53$, $p < .001$), reduced fear of infection ($r = 0.42$, $p < .001$), no more fear of legal consequences ($r = 0.40$, $p < .001$), reduced feeling of putting oneself at risk ($r = 0.35$, $p < .001$), no more fear of mistakes ($r = 0.27$, $p < .001$).

A highly significant negative correlation was found between subjective self-rating of first aid abilities (in general ($r = -0.34$, $p < .001$) and in cardiopulmonary resuscitation ($r = -0.27$, $p < .001$) and enhancement of helpfulness. This relationship shows that with telemedical support, helpfulness can be enhanced especially among persons lacking first aid skills.

Another highly significant correlation with helpfulness was found for fear of mistakes when administering first aid ($r = 0.32$, $p < .01$). Participants who showed high values for fear of mistakes also showed high values for enhancement of helpfulness with provision of a telemedical kit. Further factors significant on the $\alpha = .05$ level are previous experience of traffic accidents ($r = -0.22$, $p < .05$) and age of the potential helpers ($r = -0.21$, $p < .05$). With increasing age and with previous experience of a traffic accident, the enhancement in helpfulness with provision of telemedical guidance is lower. No significant correlations were found regarding fear of failure and attitude towards technical equipment.

4 Discussion

Three different main reasons for helpfulness could be identified: rating of own first aid ability, social obligation to help, and inhibiting fears, above all fear of mistakes. The factor first aid ability can be interpreted as subjective competency of the participants in performing first aid, since within the framework of the study, the ability to do CPR, for example, was not assessed in an objective test but with a self-rating. The parameter level of the factor social obligation to help can be seen as an indicator of feeling responsible to perform first aid.

Similar results were found by Axelsson et al. (1982) or Bierhoff et al. (1990) in interview studies. The authors identified the factors social obligation to help and competence as influential. In the decision model of Latané and Darley (1969, 1970), the crucial influencing factors can be connected with decision phase 3, realizing that one is responsible to help, and decision phase 4, deciding what kind of help is relevant (see introduction of this article).

In addition, the present study found fear of mistakes to be a prominent factor. This inhibiting factor can be related to the theory of arousal-cost-reward of helping behaviour (Piliavin et al. 1981). It is assumed that persons react in the most cost-effective manner to reduce the arousal caused by a shock. If the probability to make mistakes while helping seems to be high, a witness will leave the site of the accident or get help rather than help himself.

The present study aimed to identify factors that can be influenced through the provision of telemedical support in such a way that bystanders' willingness to perform first aid is increased. It is reasonable to assume that providing professional telemedical guidance to bystanders increases their subjective competence and reduces their fear of mistakes and infection. The feeling of responsibility, however, is a personality trait of the potential helper and cannot be influenced easily.

The results indicate that telemedical guidance increases the subjective competence of potential helpers. Bystanders whose subjective competence is rather low and who do not have a medical profession or medical education might benefit most by telemedical guidance. This assumption is further supported by the fact that no longer feeling overstrained correlates with enhanced helpfulness. The bivariate analysis supports the assumption of reduced fear of mistakes as well reduced fear of infection through the provision of telemedical support in performing first aid. The greater the individual's inherent fear of making mistakes in performing first aid, the greater the enhancement of helpfulness with the provision of a telemedical system.

With telemedical support, the anticipated costs could be reduced which results in enhancement of helpfulness. However, to have the positive effect, the affinity to technical equipment plays an important role. To ensure that a low affinity to technical equipment does not inhibit helpfulness, a high usability of the user interface should be borne in mind in the process of system development.

To interpret the results it is important to note, however, that the study is based solely on self-ratings by participants of their own behaviour in the described scenarios. Moreover, the recruitment of volunteer participants may already have led to a selection of individuals who are generally more helpful and technologically experienced than others. It also has to be kept in mind that participants might have responded according to what they believed the study's hypothesis to be. Considering this, the results of this study should be replicated or corroborated using a between-subjects design to eliminate such an effect.

5 Conclusion

Telemedical systems are nowadays being applied in all medical domains, for example as a convenient way for patients with chronic diseases living in rural areas to avoid traveling to consult a doctor or specialist. In Brandenburg, the first telemedical network in Germany is currently being tested with patients suffering heart disease (Carl-Thiem-Klinikum, 2011), which was awarded with the Jannsen Future Award 2012.

The results of the study show that the availability of a telemedical system can reduce the factors that inhibit helpfulness and therefore increase the willingness to perform first aid. For emergency situations in traffic accidents, telemedicine is not established yet. Despite all the motivational factors related to the utilization of telemedicine Rogove et al. (2012) report barriers like licensing, costs for technology, and reimbursement issues.

An alternative consists in speeding up the arrival time of the emergency team. As a fact, cars can already automatically call emergency services in case of a serious crash. This so called eCall-system is highly promoted by the European Commission and may possibly be mandatory in Europe in 2015. Emergency response times could be shortened by 40% in urban and 50% in rural areas (European Commission, 2013).

However, telemedical systems still bridge the gap between the occurrence of the accident and the arrival of the emergency team. Telemedical instruction can not only increase helpfulness in general, it can also improve the quality of first aid methods. Expert knowledge presented step-by-step to untrained helpers using a personal digital assistant (PDA) has shown to improve the quality of bystanders basic life support (Ertl & Christ, 2007). Some lives can only be saved by immediate bystander help and telemedical systems seem to be a promising way to facilitate this.

Literature

Ashton, N.L., & Severy, L.J. (1976). Arousal and Costs in Bystander Intervention. *Personality and Social Psychology Bulletin*, 2, 268-272.

- Axelsson, A., Herlitz, J. & Fridlund, B. (1982). How bystanders perceive their cardiopulmonary resuscitation intervention; a qualitative study. *Resuscitation*, 47, 71-81.
- Batson, C.D. (1998). The psychology of helping and altruism: Problems and puzzles. *Contemporary Psychology*, 43, 108-109.
- Beaman, A.L., Barnes, P.J., Klentz, B. & McQuirk, B. (1978). Increasing Helping Rates Through Information Dissemination – Teaching Pays. *Personality and Social Psychology Bulletin*, 4, 406-411.
- Bierhoff, H. W., Klein, R. & Kramp, P (1990). Hemmschwellen zur Hilfeleistung – Untersuchung der Ursachen und Empfehlung von Maßnahmen zum Abbau. Bericht zum Forschungsprojekt 8528/3 der Bundesanstalt für Straßenwesen. Bergisch Gladbach, Juli 1990.
- Billittier, A. J., Lerner, E. B., Tucker, W. & Lee, J. (1999). The Lay Public's Expectations of Prearrival Instructions when Dialing 9-1-1. *Prehospital Emergency Care*, 4, 234-237.
- Brenner, B. E., Van, D. C. Cheng, D. & Lazar, E. J. (1997). Determinants of reluctance to perform CPR among residents and applicants: The impact of experience on helping behaviour. *Resuscitation*, 35, 203-211.
- Brenner, B.E. & Kauffmann, J. (1993). Reluctance of Internists and Medical Nurses to perform Mouth-to-mouth Resuscitation. *Archives of Internal Medicine*, 153, 1763-1769.
- Carl-Thiem-Klinikum (2011). Offizieller Startschuss für brandenburgisches Telemedizin-Netz: Hochrisiko-Herzpatienten werden aus der Ferne rund um die Uhr medizinisch betreut. Retrieved July 07, 2013 from <http://www.ctk.de/Offizieller-Startschuss-f-r-brandenburgisches.2005.0.2.html>.
- Carter, W. B., Eisenberg, M. S., & Hallstrom, A. P. (1984). Development and implementation of emergency CPR instruction via telephone. *Annals of Emergency Medicine*, 13, 695-700.
- Central Union of Electrotechnology and Electrotechnical Industry e.V. (1999). Automatisches Notrufsystem -Positionspapier des ZVEI und des VDA in Zusammenarbeit mit dem DVR. Zentralverband Elektrotechnik- und Elektrotechnikindustrie e.V., Frankfurt am Main.
- Clark, R.D., & Word, L.E. (1972). Why don't bystanders help questionable – because of ambiguity. *Journal of Personality and Social Psychology*, 24, 392.
- Culley, L., Clark, J.J., Eisenberg, M.S., Larsson, M.P. (1991) Dispatcher-assisted telephone CPR: common delays and time standards for delivery. *Annals of Emergency Medicine*, 20, 362-366.
- Donner-Banzhoff, N, Schuster, C, Hofmann, S, Baum E (1999) Notfallversorgung durch Ersthelfer - Epidemiologie und Versorgungs-Qualität in einer ländli-

- chen Region. *Anästhesiologie, Intensivmedizin, Notfallmedizin, Schmerztherapie*, 34, 140-145.
- Ertl, Lorenz & Christ, Frank (2007). Significant improvement of the quality of bystander first aid using an expert system with a mobile multimedia device. *Resuscitation*, 74, 286-295.
- European Commission (2013). Pressebericht der Europäischen Kommission vom 13/06/2013. Retrieved July 25, 2013 from http://ec.europa.eu/commission_2010-2014/kallas/headlines/news/2013/06/ecall_de.htm.
- Form, W. H., & Nosow, S. (1958). *Community in Disaster* (New York: Harper).
- Forplan Dr. Schmiedel (2013). Hilfsfristen der Bundesländer (URL). Retrieved August 09, 2013 from http://www.forplan.de/tl_files/downloads/Hilfsfrist%20Bundeslaender.pdf.
- Gaertner, S. L., & Dovidio, J. F., (1977). Subtlety of white racism, arousal, and helping-behaviour. *Journal of Personality and Social Psychology*, 35, 691-707.
- Hauff, S. R., Rea, T. D., Culley, L. L., Kerry, F., Becher, L & Eisenberg, M. S. (2003): Factors impeding Dispatcher-Assisted Telephone Cardiopulmonary Resuscitation. *Annals of Emergency Medicine*, 42, 731-737.
- Huston, T., Ruggerio, M., Conner, R., & Geis, G. (1981). Bystander intervention into crime: A study based on naturally-occurring episodes. *Social Psychology Quarterly*, 44, 14-23.
- Kazdin, A. E., & Bryan, J. H. (1971). Competence and volunteering. *Journal of Experimental Social Psychology*, 7, 87.
- Kellermann, A. L. , Hackman, B. B. & Somes, G. (1989). Dispatcher-assisted cardiopulmonary resuscitation – Validation of efficacy. *Circulation*, 80, 1231-1239.
- Larsson, E. M., Martensson, N. L. & Alexanderson, K. A. E. (2002). First-aid Training and Bystander Actions at Traffic Crashes – A Population Study. *Prehospital and Disaster Medicine*, 17, 134-41.
- Latane, B. & Darley, J.M. (1969). Social determinants of bystander intervention in emergencies. In *Altruism and Helping*, J. Macaulay & L. Berkowitz (Eds.) New York: Academic Press, pp. 13-27.
- Latane, B. & Darley, J.M. (1970). *The Unresponsive bystander-Why doesn't he help?* (New York: Appleton- Century-Crofts).
- Locke, C. J., Berg, R. A., Sanders, A. B., Davis, M. F., Milander, M. M., Kern, K. B., & Ewy G. A. (1995). Bystander cardiopulmonary resuscitation. Concerns about mouth-to-mouth contact. *Archives of Internal Medicine*, 155, 938–943.
- Midlarsky, E., & Midlarsky, M. (1973). Some determinants of aiding under experimentally-induced stress. *Journal of Personality*, 41, 305-327.

- Midlarsky, E., (1971). Aiding under stress – Effects of competence, dependency, visibility, and fatalism. *Journal of Personality*, 39, 132.
- Ornato, J. P., Hallagan, L. F., McMahan, S. B., Peeples, E. H., Rostafinski, A. G. (1990). Attitudes of BCLS instructors about mouth-to-mouth resuscitation during the aids epidemic. *Annals of Emergency Medicine*, 19, 151-156.
- Pantin, H. M., & Carver, C. S. (1982). Induced competence and the bystander effect. *Journal of Applied Social Psychology*, 12, 100-111.
- Peterson, T. D. & Russel, D. W. (1999). Bystander Trauma Care Training in Iowa. *Prehospital Emergency Care*, 3, 225-230.
- Piliavin, I. M., Rodin, J., & Piliavin, J. A. (1969). Good samaritanism: an underground phenomenon? *Journal of Personality and Social Psychology*, 13, 289-299.
- Piliavin, J. A., Dovidio, J., Gaertner, S., & Clark, R. D., (1981). *Emergency Intervention*. New York: Academic Press.
- Rayko, D. S. (1977). Does knowledge matter – Psychological information and bystander helping. *Canadian Journal of Behavioural Science – Revue Canadienne des Sciences du Comportement*, 9, 295-304.
- Rogove, Herbert J., McArthur, David, Demaerschalk, Bart M. & Vespa, Paul M. (2012). Barriers to Telemedicine: Survey of Current Users in Acute Care Units. *Telemedicine & e-Health*, 18 (1), 48.
- Schlechtriemen, T., Lackner, C.K., Moecke, H., Stratmann, D. & Altemeyer, K.H. (2003). Sicherung der flächendeckenden Notfallversorgung: notwendige Strukturverbesserungen. *Notfall & Rettungsmedizin*, 6, 419-426.
- Schwartz, S., & David, A. B. (1976). Responsibility and helping in an emergency – Effects of blame, ability and denial of responsibility. *Sociometry*, 39, 406-415.
- Schwermann, T., Pape, H.C., Grotz M., Blanke, M., Greiner, W., Tröger, M., Stalp, M., Schulenburg, J.-M.G. v & Krettek, C. (2003). Einflussfaktoren auf die Überlebenswahrscheinlichkeit beim Polytrauma. *Gesundheitsökonomie & Qualitätsmanagement*, 8, 285-289.
- Shotland, R. L. & Heinold, W. D. (1985). Bystander Response to Arterial Bleeding: Helping Skills, the Decision-Making Process, and Differentiating the Helping Resonse. *Journal of Personality and Social Psychology*, 49, 347-356.
- Statistisches Bundesamt (2012). *Statistisches Jahrbuch*. Wiesbaden: Statistisches Bundesamt.
- Walz, B. (1991). Bystander Intervention – Help or Hindrance? *Journal of Emergency Medical Sciences*, 16, 60-63.
- Kidder, T. (1981). *The soul of a new machine*. Boston, MA: Little, Brown & Company.

Interdisziplinäre Systemevaluation von Ubiquitous Computing

Ludger Schmidt¹ und Kay Behrenbruch¹

Schlüsselwörter: Sozialverträglichkeit, Rechtsverträglichkeit, Gebrauchstauglichkeit, Vertrauenswürdigkeit, Sicherheit

Nutzer und Ubiquitous Computing

Mark Weiser charakterisierte Ubiquitous Computing (UC) als eine Zukunftstechnologie, die sich in die physikalische Umwelt einbindet, mit ihr verschmilzt und dem Menschen eine komplexe Unterstützung durch Informationstechnologie ermöglicht, ohne dass die "Computer" wahrgenommen werden (Weiser & Brown, 1996). UC-Systeme zeichnen sich also durch eine Einbettung der Technologie in Alltagsgegenstände und Umgebung aus, die mit dem automatischen Erfassen und Verarbeiten von Daten über den Nutzungskontext verbunden sind. UC-Systeme treffen auf der Grundlage der erfassten Daten teilweise autonome Entscheidungen, die in kontextsensitive Adaptionen umgesetzt werden. Dabei passen sich die ubiquitären Dienste in Echtzeit an den sich wandelnden Kontext an und binden sich in verschiedene technische Infrastrukturen ein. Zudem sind ubiquitäre Systeme ortsübergreifend und in Echtzeit vernetzt.

Die beschriebenen Eigenschaften von UC haben Konsequenzen für den Nutzer. Der Nutzer interagiert mit einem solchen System implizit und nicht durch ein explizites Bedienen. Auf diese Weise verbleibt die Nutzung in der Logik der durchgeführten Primäraufgabe und wird, anders als in Nicht-UC-Systemen, nicht primär von der Logik des technischen Hilfsmittels geprägt. Die Unterstützung geschieht damit einhergehend in großen Teilen außerhalb oder in der Peripherie der Aufmerksamkeit des Nutzers und steht ausschließlich dann im Aufmerksamkeitsfokus, wenn ein Bedarf gegeben ist. Der Nutzer wird also in zunehmendem Maße frei davon, seine Aufmerksamkeit und seine Handlungen an die „Bedienung“ eines Computers binden zu müssen, ohne dabei auf die IT-Unterstützung verzichten zu müssen. Mit dieser Entwicklung geht der Verlust von Kontrolle und der Möglichkeit einher, selbst zu steuern, wie sich die Anwendung verhält. Hinzu kommt, dass die erhobenen Daten sehr viel über den Nutzer aussagen und auf diese Weise ein erhebliches Potential an Verlust von Privatsphäre besteht, das von vielen Nutzer auch bewusst wahrgenommen wird.

¹ Universität Kassel, Fachgebiet Mensch-Maschine-Systemtechnik, Mönchebergstr. 7, 34125 Kassel, {L.Schmidt, K.Behrenbruch}@uni-kassel.de

Sozialverträgliche Technikgestaltung

Die Interaktion zwischen Mensch und Computer ändert sich laufend. Ein aktueller Trend wurde in dem vorangegangenen Abschnitt unter dem Schlagwort UC behandelt und die Konsequenzen für den Nutzer herausgestellt. Die nutzergerechte Gestaltung von Technik steht also immer wieder vor neuen Herausforderungen. Im Fall von UC geht es in erster Linie darum, die Produkte so zu entwickeln, dass die Herausforderungen für eine nutzergerechte Gestaltung bereits im Entwicklungsprozess adressiert und wenn möglich gut gelöst werden. Forschungsarbeiten zur Mensch-Computer-Interaktion können zu diesem Ziel beitragen. Ein Beispiel hierfür ist eine Gestaltung von UC, die bei kontextsensitiven Adaptionen eine Balance zwischen der Autonomie des Systems und der Beteiligung der Nutzer am Entscheidungsprozess erreicht, die einerseits die Vorteile von UC bewahrt und andererseits den Wunsch der Nutzer nach einem gewissen Maß an Kontrolle berücksichtigt (Evers et al., 2012). Ein anderes Beispiel ist eine Gestaltung, die es sich zum Ziel setzt, auch die Belange technikferner Nutzer in der Gestaltung von UC zu berücksichtigen (Behrenbruch et al., 2012a). Insgesamt gesehen wird aber deutlich, dass für die Gestaltung von UC Themen eine Rolle spielen, die über den disziplinären Kernbereich der Forschungen zur Mensch-Computer-Interaktion hinausgehen. Eine sozialverträgliche Gestaltung von UC braucht gebrauchstaugliche Lösungen, die gleichzeitig sicher, vertrauenswürdig und rechtsverträglich sind (Comes et al., 2012) und die neben individuellen Ansprüchen auch zu den gesellschaftlichen Normen und Gegebenheiten passen. Die sozialverträgliche Technikgestaltung zielt also darauf ab, die Herausforderungen bei der Gestaltung von UC durch die Berücksichtigung unterschiedlicher Themen anzugehen und gute Lösung für die einzelnen Nutzer und die Gesellschaft zu fördern.

Interdisziplinäre Zusammenarbeit bei der Evaluation von UC

Für eine sozialverträgliche Gestaltung von UC ist der gesamte Entwicklungsprozess entsprechend ISO 9241-210 zu berücksichtigen. Dazu gehört eine Anforderungsanalyse, die Aspekte aus den unterschiedlichen Disziplinen berücksichtigt (Hoffmann et al., 2012) sowie die Entwicklung von Konzepten und deren Implementierung. Dabei wird die Evaluation als integraler Bestandteil des Gesamtprozesses verstanden. Ziel ist es, eine sozialverträgliche Technikgestaltung zu fördern, die von Beginn an auf Rückkopplungen aus dem laufenden Entwicklungsprozess in Form von formativen Evaluationsergebnissen setzt.

Aufbauend auf den bereits diskutierten, grundlegenden methodischen Ansätzen (Schmidt et al., 2011) wird hier die konkrete Umsetzung einer interdisziplinären Evaluation dargestellt, die umfassend ist und auf der Nutzung eines reifen Prototyps mit echten Nutzern basiert. Die Evaluation wurde von einem interdisziplinären Team geplant und am Beispiel der ubiquitären Anwendung MyGroup durchgeführt.

MyGroup ist eine RFID-basierte UC-Anwendung, die Unterstützung und Erleichterung von Arbeitsabläufen im Alltag bieten soll. Alle Mitarbeiter einer Arbeitsgruppe tragen einen RFID-Tag und können über die Anwendung sehen, wo sich die anderen Kollegen aufhalten. Dadurch können sie gesuchte Kollegen schneller finden. Über die Anwendung kann jeder auch sehen, wer mit wem wie lange spricht. Durch den Einbezug des Verschlagwortungssystems BibSonomy ist es ebenfalls möglich, dass ein Mitarbeiter einen interessanten Artikel mit einem Bookmark in der Anwendung versieht. Hat auch ein anderer Mitarbeiter Informationen zu dem im Artikel behandelten Thema, wird ein Gesprächswunsch vom System vorgeschlagen. Es können auch aktiv Gesprächswünsche bei anderen Mitarbeitern angemeldet werden. Mitarbeiter können in der Anwendung auch angeben, wann sie nicht gestört werden möchten. Es besteht die Möglichkeit, mit Hilfe von Filtern und Suchfunktionen nach bestimmten Informationen in MyGroup zu suchen, z. B. nach Kompetenzen von Mitarbeitern. Die Anwendung bietet die Möglichkeit, Einstellungen zu Kontaktinformationen, zur Privatheit, zu den sozialen Netzwerken wie beispielsweise Twitter, Flickr oder Facebook, vorzunehmen. MyGroup basiert auf der Internetplattform Ubicon (Atzmüller et al., 2012).

Das methodische Rahmenkonzept für die Systemevaluation ist in Abbildung 1 dargestellt. Bei der Konzeption wurde darauf Wert gelegt, dass die Input-Output-Beziehungen auf jeder Stufe deutlich werden. Zudem wurde die Eigenständigkeit der beteiligten Disziplinen durch eine getrennte Auswertung der Ergebnisse gewahrt, ohne das Ziel zu gefährden, ein einheitliches und konsistentes Gesamtergebnis zu erreichen.

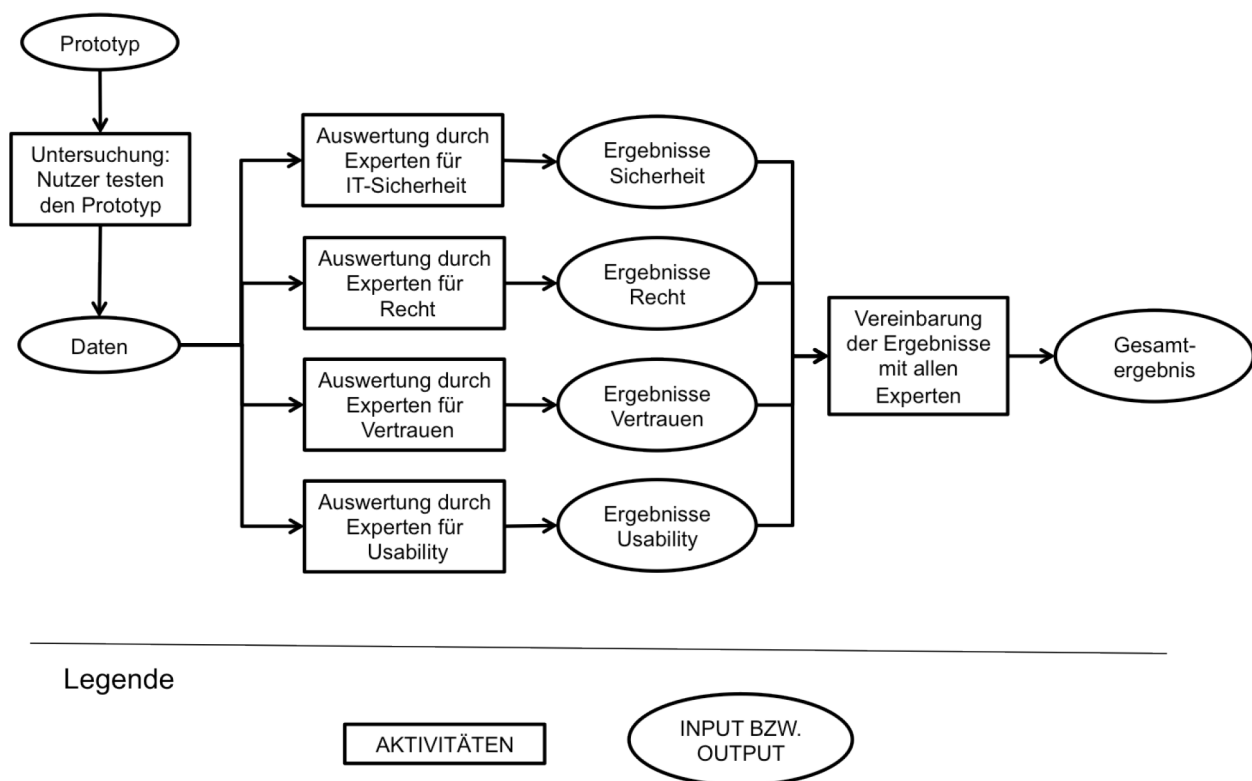


Abb.1: Methodisches Rahmenkonzept für die interdisziplinäre Evaluation

Der erste Schritt zur Integration der gemeinsamen Evaluation bestand darin, die Rahmenbedingungen der nutzerbasierten Evaluation im Vorfeld entsprechend der disziplinären Erfordernisse abzustimmen und eine Konkretisierung innerhalb des gemeinsamen Rahmenkonzepts vorzunehmen, so dass die Erhebung von für alle Disziplinen passenden Daten ermöglicht wird. Folgendes wurde festgelegt:

- Evaluation mit 20-25 Probanden, die gezielt nach Zugehörigkeit zur tatsächlichen Nutzergruppe und entsprechend der für die Akzeptanz von UC maßgeblichen Nutzereigenschaften ausgewählt wurden (Behrenbruch et al., 2012b)
- Evaluation von Funktionen in Standardsituationen und in zuvor gemeinsam entwickelten kritischen Nutzungssituationen
- Integration einer Untersuchung im Usability Labor
- Evaluationsdauer 2 Tage

Die Probanden für die interdisziplinäre Evaluation von MyGroup kamen aus den unterschiedlichen Fachgebieten des Forschungszentrums für Informationstechnik Gestaltung (ITeG) an der Universität Kassel¹, waren an dem zugrundeliegenden UC-Forschungsprojekt jedoch nicht beteiligt. Sie arbeiteten während des Experiments an ihren üblichen Arbeitsplätzen. Insgesamt waren 24 Teilnehmer in 16 Räumen in vier Stockwerken eines gemeinsamen Gebäudes involviert. Außerdem war ein Büro mit zwei Mitarbeitern an einem weiteren, zu Fuß leicht erreichbaren Standort untergebracht. Die Räume wurden mit jeweils zwei RFID-Readern ausgestattet. Für jeden Teilnehmer stand ein RFID-Tag zur Verfügung. Die Teilnehmer hatten in diesem Zusammenhang noch nicht mit einer UC-Anwendung wie MyGroup gearbeitet. Die Kommunikation zwischen ihnen erfolgte normalerweise über E-Mail und direkten persönlichen Kontakt. Bei einer vorherigen Organisationsanalyse konnte festgestellt werden, dass die Arbeitsabläufe und Arbeitsstrukturen durch MyGroup verbessert werden können. Drei Wochen vor der Systemevaluation wurden die Teilnehmer in die eingesetzte Technik und die eingesetzten Methoden eingeführt. Außerdem wurden organisatorische Hinweise zum Ablauf der Systemevaluation gegeben. Kernbestandteil der Evaluation waren die Nutzung von MyGroup im Arbeitsalltag mit der Simulation von kritischen Nutzungssituationen sowie eine Untersuchung zur Interaktion mit MyGroup im Usability Labor.

Die Auswertung der während der zweitägigen Evaluation erhobenen Daten erfolgte zunächst getrennt in den einzelnen Disziplinen (vgl. Abbildung 1) und wird in den nachfolgenden Einzelbeiträgen vorgestellt. Während der erste Beitrag "Die Simulationsstudie als Evaluationsmethode sozialverträglicher Technikgestaltung" den Evaluationsteil in der realen Anwendungssituation beschreibt, konzentrieren sich die weiteren zwei Beiträge auf den Laborteil der Evaluation. Dabei stellt der Beitrag "Gebrauchstauglichkeit im Labor objektiv evaluieren" die Erprobung einer grundlegenden Methode zur Bestimmung der Gebrauchstauglichkeit vor, die dann

¹ www.iteg.uni-kassel.de/

im Beitrag "Vertrauenswürdige Gestaltung von ubiquitären Systemen – Potentiale des Einsatzes von Methoden des NeuroIS" für die Evaluation von vertrauensunterstützenden Komponenten angewendet und bezogen auf den neuen Anwendungskontext bewertet wird. Der abschließende Beitrag "Interdisziplinäre Sicherheitsanalyse" diskutiert die Evaluation der Informationssicherheit mit Bezug zum interdisziplinären Kontext und im Hinblick auf die praktischen Erfahrungen in der gemeinsamen Evaluation.

Der sich anschließende Prozessschritt zur interdisziplinären Vereinbarung der disziplinären Ergebnisse war zum Zeitpunkt der Drucklegung noch nicht erfolgt. Die wesentlichen Schritte der in einem ganztägigen Workshop mit allen beteiligten Disziplinen durchzuführenden Abstimmung sind die folgenden:

- Definition der Darstellungsform der Ergebnisse auf einem einheitlichen Abstraktionsniveau
- Sicherstellen der Verständlichkeit aller Ergebnisse für alle Beteiligten
- Identifikation von Redundanzen und Zusammenführung von ähnlichen Ergebnissen
- Identifikation von Widersprüchen, Diskussion und Einigung auf eine Lösung mit einem gemeinsamen Ergebnis
- Ableitung von Empfehlungen für die Gestaltung zur Verbesserung des evaluierten Prototyps
- Priorisierung der Empfehlungen

Ziel dieses Vorgehens ist es, den Entwicklern redundanz- und widerspruchsfreie Ergebnisse zur Verfügung stellen zu können, die konstruktiv ausformuliert sind und deren Qualität von unterschiedlichen Fachdisziplinen intensiv diskutiert wurde.

Acknowledgement

Die vorgestellte Systemevaluation und die Entwicklung des Prototyps sind Teil des Forschungsprojekts „Gestaltung technisch-sozialer Vernetzung in situativen ubiquitären Systemen (VENUS)“, das vom Forschungszentrums für Informationstechnik-Gestaltung (ITeG) der Universität Kassel durchgeführt wird. Wir danken dem hessischen Ministerium für Wissenschaft und Kunst für die Finanzierung des Projekts im Rahmen der Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz (LOEWE). Weiterführende Informationen erhalten Sie unter: <http://www.iteg.uni-kassel.de/venus>.

Literatur

Atzmueller, M., Becker, M., Doerfel, S., Hotho, A., Kibanov, M., Macek, B. E., ... & Stumme, G. (2012, November). Ubicon: Observing Physical and Social

- Activities. In *Green Computing and Communications (GreenCom)*, 2012 *IEEE International Conference on* (pp. 317-324). IEEE.
- Behrenbruch, K., Kniewel, R., Niemczyk, S., & Schmidt, L. (2012). Akzeptanz technikferner Nutzergruppen für innovative Smartphone-Apps. In VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik (Ed.), *Useware 2012: Mensch-Maschine-Interaktion* (Kaiserslautern 2012) (Vol. 2179, pp. 65–74). Düsseldorf: VDI-Verlag.
- Behrenbruch, K., Atzmüller, M., Evers, C., Schmidt, L., Stumme, G., & Geihs, K. (2012). A Personality Based Design Approach Using Subgroup Discovery. In M. Winckler, P. Forbrig, & R. Bernhaupt (Eds.), *Human-Centered Software Engineering* (Vol. 7623, pp. 259–266). Berlin Heidelberg Springer.
- Comes, D., Evers, C., Geihs, K., Hoffmann, A., Kniewel, R., Leimeister, J., Niemczyk, S., et al. (2012). Designing Socio-technical Applications for Ubiquitous Computing. In K. Göschka & S. Haridi (Eds.), *Distributed Applications and Interoperable Systems* (Vol. 7272, pp. 194–201). Berlin Heidelberg: Springer.
- DIN EN ISO 9241-110. Ergonomie der Mensch-System-Interaktion - Grundsätze der Dialoggestaltung. (2008-09). Berlin: Beuth.
- Evers, C., Kniewel, R., Geihs, K., & Schmidt, L. (2012). Achieving User Participation for Adaptive Applications. In J. Bravo, D. López-de Ipina, & F. Moya (Eds.), *Ubiquitous Computing and Ambient Intelligence: 6th International Conference* (Vitoria-Gasteiz 2012) (Vol. 7656, pp. 200–207). Berlin: Springer.
- Hoffmann, A., Hoffmann, H. & Leimeister, J. M. (2012). Anforderungen an Software Requirement Pattern in der Entwicklung sozio-technischer Systeme. In U. Goltz, M. A. Magnor, H.-J. Appelrath, H. K. Matthies, W.-T. Balke & L. C. Wolf (eds.), *42. Jahrestagung der Gesellschaft für Informatik e.V.* (pp. 379-393). Braunschweig: GI.
- Schmidt, L., Leimeister, J. M., & Behrenbruch, K. (2011). Methoden-Evolution für die Entwicklung ubiquitärer Anwendungen. In S. Schmid, M. Elepfandt, J. Adenauer, & A. Lichtenstein, A. (Hrsg.): *Reflexionen und Visionen der Mensch-Maschine-Interaktion - Aus der Vergangenheit lernen, Zukunft gestalten: 9. Berliner Werkstatt Mensch-Maschine-Systeme* (Berlin 2011) (pp. 186–187). Düsseldorf: VDI-Verlag.

Die Simulationsstudie als Evaluationsmethode sozial-verträglicher Technikgestaltung

Alexander Roßnagel¹ und Michaela Schuldt¹

Schlüsselwörter: Simulationsstudie, interdisziplinär, Evaluation, sozialverträgliche Technikgestaltung

Sozialverträgliche Technikgestaltung

Die IT-Systeme der nächsten Generation werden mit Sensoren den jeweiligen Anwendungskontext erfassen und spontan mit anderen ubiquitären Diensten in ihrer Umgebung interagieren. Es entsteht damit eine ganz neue Art der vernetzten Informationsverarbeitung. Für ihre Nutzung in der Praxis benötigt die neue Technik eine Einbettung in die Gesellschaft. Da sie ihre Leistungen überall, jederzeit und vernetzt erbringen soll, wird sie damit eine starke Wirkung auf die soziale Vernetzung und Vergesellschaftung ausüben. Sie muss den Zielen der Gesellschaft, die diese sich für ihr Zusammenleben selbst gesetzt hat, entsprechen. Dass diese Einbettungen gelingen, ist ein entscheidender Erfolgsfaktor für die neue Technik.

Dies wird aber nur der Fall sein, wenn die UC-Anwendungen von Anfang an so gestaltet sind, dass sie sozialverträglich, also mit der gesellschaftlichen Ordnung und Entwicklung vereinbar sind. Das setzt u.a. anderem voraus, dass UC-Anwendungen die Selbstbestimmung der Nutzer und andere Zielsetzungen des Rechts berücksichtigen, die sich die Gesellschaft in einem formellen demokratischen Verfahren selbst gesetzt hat. Um dieses Ziel der sozial- und rechtsverträglichen Gestaltung von UC-Anwendungen zu erreichen, hat das Forschungsprojekt „Gestaltung technisch-sozialer Vernetzung in situativen ubiquitären Systemen“ (VENUS) eine Methode entwickelt, mit der aus normativen Vorgaben technische Gestaltungsvorschläge erarbeitet und evaluiert werden können. Eine für UC-Anwendungen besonders geeignete Evaluationsmethode ist die „Simulationsstudie“, die im Folgenden näher vorgestellt werden soll.

Die Methode der Simulationsstudie

Die Simulationsstudie ist eine Evaluationsmethode, die es Anwendern erlaubt, spezifische Erfahrungen mit einer neuen Technik zu gewinnen, ohne dabei Risiken einzugehen. Sie ist besonders geeignet, wenn es nicht nur um das Verhältnis von Mensch und Maschine, sondern um Technik geht, die in der vernetzten Interaktion zwischen Menschen eingesetzt wird. Mit ihr soll das sozio-technische System anhand von Kriterien der Sozial- und Rechtsverträglichkeit innerhalb der Interaktion untersucht werden. Aus den Erfahrungen der Anwender können Vorschläge zur

¹ Universität Kassel, Fachgebiet Öffentliches Recht, Pfannkuchstr. 1, 34121 Kassel, {A.Rossnagel, M.Schuldt}@uni-kassel.de

Verbesserung der Technik abgeleitet werden (siehe zur Methode der Simulationsstudie: Roßnagel et al., 1994; Roßnagel et al., 1999; Roßnagel, 2002; Fischer-Dieskau et al., 2004; Fischer-Dieskau et al., in: Roßnagel & Schmücker 2006; Gitter et al., 2007; Roßnagel et al. 2009).

Viele Fehler und Verbesserungsmöglichkeiten eines Techniksystems zeigen sich häufig erst bei seiner Anwendung. In ihrer vollen Tragweite werden sie erst in der praktischen Erprobung erkennbar. Jedoch können Gestaltungsvorschläge zur Behebung dieser Fehler oft nur berücksichtigt werden, bevor die Entwicklungsprozesse abgeschlossen, die finanziellen Investitionen getätigt, die organisatorischen Prozesse umgestellt und rechtliche Beurteilungen vorgenommen worden sind. Ist eine technische Entwicklung erst einmal abgeschlossen, besteht ein Sachzwang, sich der Technik unverändert zu bedienen. Deshalb müssen Bewertung und Gestaltung von Technik vorausschauend erfolgen. Dabei kann insbesondere die Erprobung nicht nur durch die Entwickler, sondern auch durch die Anwender wichtige Erkenntnisse liefern.

Ziel der Erprobung ist mithin, im Interesse einer frühzeitigen, präventiven Technikgestaltung Erfahrungen zu gewinnen. Dies führt jedoch auf vielen Anwendungsgebieten zu Widersprüchen, denn die Technik, mit der Erfahrungen gesammelt werden sollen, liegt lediglich prototypisch vor. Die potentiellen Anwender sind auf den Umgang mit der Technik noch nicht vorbereitet und das erforderliche Vorwissen fehlt ihnen. Überdies sind die rechtlichen Rahmenbedingungen für die fragliche Technik noch nicht umfassend festgelegt. Darüber hinaus sollen Erfahrungen mit Risiken und Folgen gesammelt werden, die in der Wirklichkeit gerade vermieden werden müssen. Aber gerade deshalb müssen auch Extremsituationen erprobt werden. Je größer die Risiken und je schwerwiegender die Folgen sind, umso mehr sind Erfahrungen erforderlich, umso weniger aber sind die zum Gewinn dieser Erfahrungen notwendigen Maßnahmen – zumal nur für Forschungszwecke – zu verantworten (Roßnagel et al., 2009).

Die Simulationsstudie ist eine Methode, die in der Lage ist, mit den dargestellten Paradoxien umzugehen: Zum einen soll mit einer Technik Erfahrung gesammelt werden, die es noch nicht gibt, und zum anderen sollen Erfahrungen mit Risiken und Folgen gemacht werden, die zu verhindern sind. Ziel der Simulationsstudie ist es, systematisch Erfahrungen zu sammeln nach dem Prinzip „höchstmögliche Realitätsnähe unter Vermeidung von Schäden“ (Roßnagel & Sarbinowski, 1993; Pordesch et al., 1993; Roßnagel et al., 1998). Um dieses Ziel zu erreichen ist es notwendig,

- mit echten Techniksystemen Erfahrungen zu sammeln, die prototypisch entwickelt sind,
- mit echten Anwendern zu arbeiten, die als „sachverständige Testpersonen“ mit ihren Erfahrungen die Technik besser beurteilen können,
- mit echten Problemen umzugehen, die realen Fällen nachgestellt sind,

- mit echtem Arbeitsmaterial zu arbeiten, das ausschließlich hierfür vorbereitet wurde,
- mit echten Angriffen und Pannen Erfahrungen zu sammeln, wobei die Folgen nur innerhalb der Simulation auftreten und nicht real sind,
- mit echten Testfällen zu arbeiten, die ebenfalls keine realen Schäden verursachen (Roßnagel & Sarbinowski, 1993; Pordesch et al., 1993).

Die Simulationsstudie übernimmt verschiedene Aspekte von anderen Methoden, modifiziert diese aber, um bestimmte Nachteile zu vermeiden. Zunächst übernimmt sie den Gedanken einer Simulation, die einen Blick in die Zukunft gewährt. Bei einer Simulation wird jedoch kein soziales Feld abgebildet. Dies bietet die Simulationsstudie, indem sie mit echten Anwendern arbeitet, die das Techniksystem als sachverständige Testpersonen überprüfen (Roßnagel et al., 1998). Die Simulationsstudie adaptiert Aspekte der partizipativen Technikgestaltung. Sie bindet Personen ein, die die Technik, mit ihrem Organisationswissen und ihrer Arbeitserfahrung nutzen sollen. Ein Nachteil partizipativer Technikgestaltung ist jedoch, dass die Teilnehmer die zukünftigen Nutzungssituationen nicht abschätzen können. Dieser Nachteil wird von der Simulationsstudie beseitigt, indem sie zukünftige Nutzungssituationen ermöglicht. Außerdem werden Aspekte von Feldversuchen in der Simulationsstudie integriert. Ein Feldversuch bezieht echte Anwender über einen längeren Zeitraum im Routinebetrieb ein. Dadurch können langfristige Effekte erkannt werden. Bei der Simulationsstudie werden ebenfalls echte Anwender einbezogen. Jedoch wird durch gezielte Fallkonstellationen spezifische Erfahrung über einen kürzeren Zeitraum erzeugt. Dadurch werden hohe Kosten beim Veranstalter und ein erhöhter Aufwand bei den Teilnehmern vermieden. Durch den kürzeren Einsatzzeitraum gelangt man schneller zu Ergebnissen. Darüber hinaus können in einer Simulationsstudie spezifische, insbesondere kritische Nutzungssituationen gezielt mehrfach untersucht werden, die in einem Feldversuch niemals oder sehr selten auftreten.

Um systematisch Erfahrungen sammeln zu können und später zu anwendernahen Gestaltungsvorschlägen zu kommen, erfolgt die Simulationsstudie in verschiedenen Schritten (Roßnagel et al., 2009; Roßnagel & Sarbinowski, 1993; Pordesch et al., 1993; Roßnagel et al. 1998).

Simulationsstudie VENUS

Die Simulation VENUS wurde am 17. und 18. Juli 2013 im Forschungszentrum für Informationstechnik-Gestaltung (ITeG) in Kassel durchgeführt. Evaluiert wurde dabei die UC-Anwendung MyGroup.

Ausreichend vor der Simulation wurden Fälle konstruiert, um sowohl Standardfunktionen der Technik als auch kritische Nutzungssituationen zu testen. Die Fälle orientierten sich an den Evaluationskriterien der Sozial- und Rechtsverträglichkeit. Anhand der Fälle wurde die Umsetzung der zuvor mit der VENUS-Methode erar-

beiteten Gestaltungsvorschläge überprüft. Bei der Fallerarbeitung wurde sichergestellt, dass die Arbeitsaufgaben einigermaßen gleich auf die Teilnehmer und die Zeit verteilt wurden. Auch wurde festgehalten, wie die vorbereiteten Aufgaben der jeweiligen sachverständigen Testperson mitgeteilt (E-Mail oder Karte) und wie deren Bearbeitung beobachtet und ausgewertet werden sollte.

Insgesamt wurden 81 Fälle erarbeitet. Sie wurden zu einem „Drehbuch“ zusammengefügt. Zunächst wurde eine zeitliche Reihenfolge festgelegt. Danach wurden die einzelnen Aufgaben den Teilnehmern zugeordnet. Außerdem fand eine Zuordnung der einzelnen Fälle zu Personen der Simulationsleitung statt. Nach der Personenzuordnung erfolgte eine Überprüfung von Kollisionen in Bezug auf den Zeitablauf.

Im Vorfeld mussten die Ereigniskarten und E-Mails vorbereitet werden. Die Technik wurde eine Woche vor der Simulation installiert und intensiv durch die Simulationsleitung mit Unterstützung der Entwickler getestet. Dabei wurden schon Daten generiert, um manche Fälle vorzubereiten. Außerdem wurde eine datenschutzrechtliche Einwilligungserklärung verfasst, die von jedem Teilnehmer unterzeichnet wurde.

Am ersten Tag der Simulation verteilte die Simulationsleitung die RFID-Tags an die Teilnehmer. Diese gingen dann an ihren Arbeitsplätzen ihrer gewohnten Arbeit nach. Im Laufe der nächsten sechs Stunden wurden den Teilnehmern durch die Simulationsleitung die Fälle gegeben oder zugesendet. Im Anschluss an die zu erledigenden Aufgaben aus den Fällen wurden die Testpersonen von einem Mitglied der Simulationsleitung zu ihren Erfahrungen befragt. Die Simulationsleitung war während des ganzen Tages für eventuelle Rückfragen erreichbar. Zur Simulationsleitung gehörte auch ein Systemadministrator, der eventuell auftretende technische Probleme lösen sollte. Am zweiten Tag wurde die Simulation ebenfalls über sechs Stunden durchgeführt.

Zur Auswertung wurden verschiedene Methoden angewandt, um die unterschiedlichen Erfahrungen der sachverständigen Testpersonen den Auswertern zugänglich zu machen. Diese reichen von der teilnehmenden Beobachtung und deren Protokollierung über Eigenprotokolle der Testpersonen, Einzel- und Gruppeninterviews bis hin zu einem Online-Fragebogen. An beiden Tagen wurde jeweils ein Abschlussgespräch durchgeführt, das die jeweils einzelnen Erfahrungen der Testpersonen abglich und bestimmte Gestaltungsaspekte fokussierte.

Ergebnisse

Die Simulationsstudie hat in Bezug auf die Fragen der Sozial- und Rechtsverträglichkeit relevante Ergebnisse erbracht. Es konnten neue und verbesserte Gestaltungsvorschläge für das evaluierte System erarbeitet werden. In diesem Beitrag werden jedoch nur exemplarisch die Ergebnisse der Kriterien der Transparenz, der Erforderlichkeit, der Datenvermeidung und Datensparsamkeit sowie der Werkzeugeignung vorgestellt.

Die Erfüllung des Kriteriums der Transparenz setzt voraus, dass MyGroup so gestaltet ist, dass die Nutzer in der Lage sind, zu erfahren, wer was wann und bei welcher Gelegenheit über sie weiß und wie das System funktioniert. Durch eine transparente Gestaltung von MyGroup können etwa die informationelle und die kommunikative Selbstbestimmung gefördert werden. Umfasst wird von diesem Kriterium sowohl eine datenschutzrechtliche als auch eine funktionsbezogene Transparenz. Letztlich sind zur Umsetzung des Kriteriums der Transparenz eine übersichtliche Gestaltung, die Klarheit der Funktionen und Kenntnismöglichkeit der verarbeiteten personenbezogenen Daten sowie die Nachvollziehbarkeit der Prozesse erforderlich.

Die Simulationsstudie hat gezeigt, dass die Informationen in den Pop-Ups als positiv empfunden wurden, um zur Transparenz des Systems beizutragen. Jedoch gibt es Funktionen des Systems, die von den sachverständigen Testpersonen nicht verstanden wurden. Beispielsweise war die jeweilige Bedeutung von Einstellungen zur Privatheit und zum Status unklar. Dies kann durch weitere Informationen, die die Funktionen erklären, beseitigt werden. Außerdem ist die Gestaltung der Bedienoberfläche unübersichtlich. Die sachverständigen Testpersonen haben einige Funktionen nicht gefunden oder auch interaktive Elemente übersehen. Das System sollte daher so gestaltet werden, dass mit Farbenbedeutungen gearbeitet wird und eine auffälligere Gestaltung gegeben ist. Die sachverständigen Testpersonen wussten auch nicht, wann eine Aktion des Systems ausgeführt wurde. Eine Rückmeldung vom System, ob eine Funktion Erfolg hatte oder nicht, wäre wünschenswert. Darüber hinaus sollte es auch eine Möglichkeit zur Überprüfung geben, die das eigene Profil aus der Sicht eines anderen Nutzers anzeigt, um Einstellungen bezüglich der Privatheit kontrollieren zu können. Ebenfalls bemängelt wurde die nicht einheitliche und eindeutige Wortwahl. Dies kann durch eine Vereinheitlichung beseitigt werden. Außerdem sollte das System so gestaltet sein, dass klare Bezeichnungen gewählt werden. Den sachverständigen Testpersonen war nicht bewusst, wer was wann wusste. Dies kann durch mehr Erläuterungen zu Sinn und Zweck einzelner Funktionen des Systems verhindert werden. Beispielsweise könnte es ein Handbuch oder eine Hilfe-Seite geben. Das System könnte aber auch so gestaltet sein, dass die Bedienung sich von selbst ergibt. Des Weiteren war den sachverständigen Testpersonen nicht deutlich, was der Arbeitgeber wusste und wer Rechte als Administrator besaß. Dieses Problem kann durch Hinweise für den Nutzer gelöst werden – etwa durch "Frequently Asked Questions".

Nach dem Kriterium der Erforderlichkeit soll MyGroup Beeinträchtigungen auf das Maß beschränken, das zur Erfüllung der Aufgaben, die mit der Anwendung verfolgt werden, und zur Erreichung der jeweiligen konkreten Zwecke erforderlich ist. Nur so kann sichergestellt werden, dass ein unbefangener Umgang mit dem System die informationelle und die kommunikative Selbstbestimmung sowie eine autonome Arbeitsgestaltung nicht über Gebühr eingeschränkt.

In der Simulationsstudie zeigte sich, dass die Arbeitgeber sich ein gutes Bild von ihren Mitarbeitern machen konnten. Berufliche Daten aus sozialen Netzwerken, die

mit MyGroup verknüpft sind, können für den Arbeitgeber hilfreich sein. Mit ihnen kann er sich beispielsweise über besondere Fähigkeiten und Kenntnisse seiner Mitarbeiter informieren. Für die Arbeitnehmer ist jedoch diese Möglichkeit von MyGroup nicht klar gewesen. Dies sollte daher auf in geeigneter Weise von MyGroup erklärt werden. Außerdem fanden es die Nutzer nicht notwendig, dass der Vorname mit mindestens drei Zeichen angegeben werden musste. Der Nachname ist im beruflichen Umfeld ausreichend. Nur bei Verwechselungsgefahr sollte die Option bestehen, den Vornamen anzugeben. Weiter wurde es nicht für erforderlich gehalten, dass der Standort angegeben ist. Es würde ausreichen, dass Gesprächswünsche angegeben werden können. Außerdem sollte klar zwischen privaten und beruflichen Daten getrennt werden. Dies spielt auch bei der Zweckbindung eine Rolle. Es sollte ermöglicht werden, dass MyGroup auf die berufliche Nutzung beschränkt werden kann.

Das Kriterium der Datenvermeidung und Datensparsamkeit verlangt, Datenverarbeitungssysteme so zu gestalten, dass im Rahmen des Möglichen auf den Umgang mit personenbezogenen Daten vollständig verzichtet oder auf ein Minimum beschränkt wird. Dem Kriterium lässt sich außerdem eine zeitliche Beschränkung entnehmen. Personenbezogene Daten sind danach möglichst frühzeitig zu löschen, zu anonymisieren oder zu pseudonymisieren. Auch wenn bei MyGroup ein Verzicht auf den Umgang mit personenbezogenen Daten nicht denkbar ist, sollte die Anwendung zumindest so gestaltet sein, dass – soweit möglich – auf Dateneingaben und Datenverarbeitungen verzichtet und – soweit das den Aufgaben des Systems nicht widerspricht – anonymes und pseudonymes Handeln ermöglicht wird.

Positiv wurde festgehalten, dass MyGroup die Standortdaten nur im aktuellen Zustand anzeigt und nicht die letzten Aufenthaltsorte. Außerdem werden nur die letzten drei Bookmarks, Publikationen oder Kontakte angezeigt. Auch sollte der Nutzer steuern können, welche Daten aus den sozialen Netzwerken genutzt, wann Standortdaten erhoben und wann der Account gelöscht werden kann. Problematisch im Zusammenhang mit diesem Kriterium ist, dass die Nützlichkeit des Systems darunter leiden könnte, je weniger Daten man angibt. Aufgrund dessen sollte ein Gleichgewicht zwischen den Nutzern bestehen, wie viele Daten angegeben werden. Dies könnte eventuell durch Leitlinien bei Einführung des Systems gelöst werden.

MyGroup soll Unterstützung bei der Erfüllung von Aufgaben im Arbeitsumfeld bieten. Damit die Anwendung einer autonomen Arbeitsgestaltung einzelner Mitarbeiter dienen kann, muss es das Kriterium der Werkzeugeignung erfüllen. Das Kriterium der Werkzeugeignung setzt voraus, dass der Nutzer durch die Anwendung in seiner Arbeit unterstützt und ihm die Arbeit erleichtert wird. Dabei sind Überschneidungen mit den rechtlichen Kriterien der Transparenz und der Entscheidungsfreiheit denkbar.

Die Simulationsstudie hat ergeben, dass MyGroup viele Funktionen hat, die als vorteilhaft angesehen werden. Zum einen empfanden die sachverständigen Testpersonen es als positiv, dass man die Kollegen leichter finden konnte und auch

Forscherkollegen schneller zu identifizieren waren. Die Bedienung war einfach. Die Statusfunktion ist vorteilhaft, um Arbeiten ungestört erledigen zu können. MyGroup hilft den Mitarbeitern auch beim Herausfinden von Expertisen und dem Sammeln von Kontaktdaten. Die Lokalisierung kann bei verteilt sitzenden Arbeitsgruppen sehr hilfreich sein. Für kleinere Unternehmen kann MyGroup auch als Personenverzeichnis sinnvoll sein. Bei der Lokalisierung war jedoch die „Map“-Ansicht nicht eindeutig. Das System sollte einen richtigen Raumplan und eine Funktion, Räume zu buchen, zur Verfügung stellen. Außerdem haben die sachverständigen Testpersonen sich eine Kommunikationsmöglichkeit innerhalb von MyGroup gewünscht. Diese sollte durch einen Instant Messenger oder ähnliches geschaffen werden. Dadurch können – entsprechend dem Wunsch der Testpersonen – auch größere Treffen einberufen werden. Die Tag-Cloud hat sich ausschließlich aus Tags des Verschlagwortungssystems BibSonomy-zusammengesetzt. Dies schließt nicht aus, dass auch unpassende Kompetenzen mit einem Profil in Verbindung gebracht werden. Die Tag-Cloud sollte sich aus allen Bezugsquellen von MyGroup zusammensetzen. Die sachverständigen Testpersonen haben sich außerdem die Einbindung von Kalender und E-Mail gewünscht.

Fazit

Zur Evaluation der Sozialverträglichkeit von UC-Anwendungen bedarf es einer Methode, die sowohl den Anforderungen der Sozialverträglichkeit als auch dem spezifischen Charakter der technisch-sozialen Vernetzung der Nutzer in situativen ubiquitären Systemen gerecht wird. Sie sollte also die künftigen Nutzer einbinden und deren vernetzte Interaktion zu Gegenstand der Evaluation machen. Hierfür reichen reine Techniktests nicht aus. UC-Systeme sind sozio-technische Systeme, die ihre Nutzer sozial vernetzen. Sie können nur angewendet werden, wenn die Technik in Handlungsabläufe und gesellschaftliche Strukturen eingebettet wird. Die Entwicklung sozio-technischer Innovationen bedarf daher auch einer Evaluation, die es erlaubt nicht nur die technischen, sondern darüber hinausgehend die sozio-technischen Aspekte zu beleuchten. Dies erfordert die Einbindung der Nutzer und die Erprobung der Technik in „sozialer Aktion“. Hierfür reichen aber auch einzeldisziplinäre Evaluationen nicht aus. Für die integrierende Bewertung einer Technik-Anwendung aus dem Blickwinkel der Sozialverträglichkeit ist es gerade notwendig, mögliche Wechselwirkungen und Widersprüche unterschiedlicher Gestaltungsvorschläge in der ganzheitlichen Anwendung der Technik zu erkennen. Für eine solche Evaluation ist die Simulationsstudie eine geeignete Methode.

Die Simulationsstudie ist Teil des Entwicklungsprozesses und liefert Bewertungen und Verbesserungsvorschläge für eine prototypische Implementierung. Hierfür genügen qualitative Erprobungen. Sie zielen auf Erfahrungen mit bestimmten Nutzungssituationen, die gezielt herbeigeführt werden. Mengenrepräsentative abschließende Bewertungen sind diesem Entwicklungsstadium unangemessen. Im Gegensatz zu Prototypen würden sie eine fertig entwickelte Technik voraussetzen. Eine solche Bewertung ist nicht das Ziel einer Simulationsstudie. Mit ihr können

auch keine langfristigen Gewöhnungseffekte erkannt werden. Hierfür wären Feldstudien erforderlich.

Die Simulationsstudie bezieht die Nutzer ein, für die eine sozialverträgliche Gestaltung der Technikanwendungen Vorteile bringen soll. Ob dies der Fall ist, können diese selbst erfahren. Sie führt zu einer Erprobung der Technik-Anwendung in der vernetzten Interaktivität der Nutzer. Sie müssen die Anwendung in ihre professionellen Handlungsabläufe integrieren und diese auf die Technik abstellen. Dies ermöglicht Erfahrung mit einer gelingenden oder widersprüchlichen Einbettung der Technik. Die Simulationsstudie bietet insoweit die Möglichkeit, die für die Gestaltung der Technik und ihres organisatorischen Rahmens erforderlichen Erfahrungen in einem geschützten und kontrollierten Umfeld zu machen und dabei auch riskante Situationen zu erleben, ohne Schaden zu nehmen. Dadurch können auch gefahrlos zuvor nicht erkannte Risiken identifiziert werden.

Die Simulationsstudie bietet einen erfahrungsgestützten Gedankenaustausch zwischen normativen Disziplinen, Entwicklern und Nutzern. Dadurch können Bewertungen verglichen, Verbesserungsmöglichkeiten geprüft und neue Gestaltungsvorschläge gesammelt und diskutiert werden. Die Simulationsstudie ersetzt nicht die Evaluationen im Labor oder im Techniktest. Sie ergänzt diese aber um eine für die sozial- und rechtsverträgliche Technikgestaltung entscheidende Dimension, nämlich die realistische Nutzung der Technik durch professionelle Nutzer in ihrer sozialen Interaktion.

Literatur

- Fischer-Dieskau, S., Roßnagel, A., Steidle, R. (2004), Beweisführung am seidenen Bit-String? – Die Langzeitaufbewahrung elektronischer Signaturen auf dem Prüfstand, MMR, 451-455.
- Fischer-Dieskau, S., Pordesch, U., Roßnagel, A., Steidle, R. (2006), Simulationsstudie, in: Roßnagel, A./Schmücker, P. (Hrsg.), *Beweiskräftige elektronische Archivierung - Bieten elektronische Signaturen Rechtssicherheit?*, Bonn: Economica.
- Gitter, R., Lotz, V., Pinsdorf, U., Roßnagel, A. (2007), *Sicherheit und Rechtsverbindlichkeit mobiler Agenten*, Wiesbaden: Vieweg+Teubner Verlag.
- Pordesch, U., Roßnagel, A., Schneider, M. (1993), Erprobung sicherheits- und datenschutzrelevanter Informationstechniken mit Simulationsstudien, DuD, 491-497.
- Roßnagel, A. (2002), *Datenschutz beim Online-Einkauf – Herausforderungen, Konzepte, Lösungen*, Braunschweig: Vieweg Verlag.
- Roßnagel, A., Haux, R., Herzog, W. (1998), *Mobile und sichere Kommunikation im Gesundheitswesen*, Braunschweig: Vieweg+Teubner Verlag.

- Roßnagel, A., Sarbinowski, H. (1993), Simulationsstudien zur Gestaltung von Telekooperationstechnik – Wir brauch(t)en eine neue Methode, *GMD-Spiegel*, 30-34.
- Roßnagel, A., Schmidt, A., Wilke, D. (2009), *Rechtssichere Transformation signierter Dokumente – Anforderungen, Konzepte, Umsetzung*, Baden-Baden: Nomos.
- Roßnagel, A. u.a. (1994), *Die Simulationsstudie Rechtspflege - Eine neue Methode zur Technikgestaltung für Telekooperation*, Berlin: Ed. Sigma.
- Schulz, T., Skistims, H., Zirfas, J., Atzmüller, M., Scholz, C. (2013), Rechtliche Ausgestaltung sozialer Konferenzplattformen – Technische Lösungsansätze für datenschutzrechtliche Probleme, *ZD*, 60-65.

Gebrauchstauglichkeit im Labor objektiv evaluieren

Kay Behrenbruch¹ und Ludger Schmidt¹

Schlüsselwörter: Videobasierte Emotionserkennung, physiologische Messungen, elektrodermale Aktivität, Eye-Tracking

Zusammenfassung

Dieser Beitrag beschreibt die Ergebnisse von zwei Experimenten im Usability Labor unter Einsatz von Eye-Tracking, videobasierter Emotionserkennung bezogen auf den Gesichtsausdruck sowie der Messung der elektrodermalen Aktivität. Die Auswertung der erhobenen Daten konzentriert sich auf die Integration der drei unterschiedlichen Datenquellen und die Bildung von aussagekräftigen Kennwerten im Hinblick auf die Beurteilung der Gebrauchstauglichkeit von graphischen Benutzungsoberflächen. Im Rahmen des ersten Experimentes wird der Nachweis geführt, dass die gebildeten Kennwerte die Gebrauchstauglichkeit einer Benutzungsoberfläche abbilden und die Identifikation von kritischen Interaktionselementen ermöglichen. Das zweite Experiment untersucht die Anwendbarkeit dieser Ergebnisse im Rahmen der Evaluation der ubiquitären Software MyGroup.

Einleitung

In vielen Usability-Laboren finden sich Möglichkeiten zur Bestimmung von physiologischen Messgrößen. Es kommt allerdings auch häufiger vor, dass dieser „biometrische“ Teil der Labore in der Praxis wenig genutzt wird. Ein Grund hierfür liegt darin, dass die korrekte Interpretation von Daten über die Physiologie von Probanden eine große Herausforderung darstellt (Schmidtke, 2002). Ein Anstieg der Herzfrequenz ist beispielsweise ein viel indirekterer Indikator für die Qualität einer Software als die wertende Aussage einer Testperson. Dieser Beitrag entwirft daher zunächst am Beispiel der elektrodermalen Aktivität (EDA) einen Interpretationsrahmen für die Auswertung von physiologischen Daten bei der Evaluation von grafischen Interaktionselementen. Dazu wird die Messung der EDA mit Eye-Tracking und einer mimikbasierten Emotionserkennung kombiniert.

Emotionen werden häufig in den Dimensionen Valenz und Arousal (Schmidt-Atzert, 1996) beschrieben. Für die Beurteilung einer evaluierten Software ist es also wichtig festzustellen, ob die Reaktion des Probanden positiv oder negativ ist (Valenz) und wie stark die Reaktion ausgeprägt ist (Arousal). Für die Erfassung von Valenz und Arousal wurde im Rahmen der vorliegenden Arbeiten bei der Nutzung von Software eine videobasierte, automatische und softwaregestützte Emotionserkennung bezogen auf den Gesichtsausdruck nach Ekman (1970) verwendet. Bezogen auf die emotionalen Erregung (Arousal) wurde parallel die EDA erfasst.

¹ Universität Kassel, Fachgebiet Mensch-Maschine-Systemtechnik, Mönchebergstr. 7, 34125 Kassel, {L.Schmidt, K.Behrenbruch}@uni-kassel.de

Im Folgenden wird die Bildung von zwei Kennwerten beschrieben, die Valenz und Arousal integrieren und zu jedem Zeitpunkt der Evaluation und zusammenfassend für definierte Zeitabschnitte die emotionale Ausprägung der Probanden quantitativ beschreiben. Während der erste Kennwert sich ausschließlich auf die Daten aus der Mimikererkennung stützt, integriert der zweite Kennwert die physiologischen Daten mit den Daten aus der Mimikanalyse. Durch das ebenfalls parallel durchgeführte Eye-Tracking und die Definition von mehreren „Area of Interest“ (Schmidt, 2009) wird es möglich, die Zuordnung der Kennwerte zu einzelnen Interaktionselementen vorzunehmen.

Zudem werden Hypothesen zum Zusammenhang zwischen der Gebrauchstauglichkeit und den ermittelten Kennwerten aufgestellt und im Rahmen eines ersten Laborexperiments überprüft. In einem zweiten Laborexperiment wird die Anwendbarkeit der erzielten Ergebnisse im Rahmen der Evaluation der Gebrauchstauglichkeit einer ubiquitären Software untersucht. Abschließend wird ein Fazit gezogen und ein Ausblick gegeben.

Kennwerte

Die im Rahmen der Experimente verwendete Mimikanalyse liefert für sechs unterschiedliche emotionale Reaktionen Werte zwischen 0 und 100, die die Stärke jeder der Emotionen (Arousal) synchron zur Experimentalaufgabenbearbeitung beschreibt. Anhand dieser Werte wurde für jeden Zeitpunkt die vorherrschende emotionale Reaktion bestimmt. Für die Bestimmung der Valenz aus den Daten der Mimikanalyse wurden die ermittelten vorherrschenden Emotionen in zwei Gruppen eingeteilt:

- Für die Bewertung der Gebrauchstauglichkeit kritische emotionale Reaktionen: ärgerlich, überrascht, traurig, angeekelt (Valenz negativ)
- Für die Bewertung der Gebrauchstauglichkeit unkritische emotionale Reaktionen: neutral, freudig (Valenz positiv)

Die Einordnung der an sich nicht eindeutig negativen oder positiven emotionalen Reaktion "überrascht" als kritisch für die Gebrauchstauglichkeit ist damit begründet, dass bei der Interaktion mit einer Software, die nicht aus dem Unterhaltungs- oder Spielebereich stammt, Überraschungen im Normalfall auf Brüche der bestehenden Erwartungshaltung an die Funktionalität zu bewerten sind und damit einer guten Gebrauchstauglichkeit entgegenstehen. Die emotionalen Reaktionen "traurig" und "angeekelt" werden berücksichtigt und nicht von der Analyse ausgeschlossen, weil zu beobachten war, dass Irritationen bei der Interaktion mit der Software häufig einer dieser Kategorien zugeordnet wurden. Die Reaktionen "neutral" und "freudig" werden trotz der unterschiedlichen Wertigkeit in einer Kategorie zusammengefasst, weil im Hinblick auf die Fragestellung bei beiden Emotionen von einer guten Gebrauchstauglichkeit ausgegangen werden kann.

Der Kennwert "Mimik" wird aus dem Wert für die Stärke der vorherrschenden Emotion (Arousal) und einem Vorzeichen entsprechend der Valenz gebildet. Der Kennwert "Mimik" ist also positiv, wenn Ärger oder Freude vorherrschen (positive Valenz). Herrscht eine der mit negativer Valenz verbundene Emotion (z.B. Ärger) vor, so erhält der Kennwert ein negatives Vorzeichen.

Der zweite Kennwert "Physio" berücksichtigt die EDA-Daten zur Bestimmung des Arousals. Dazu werden die EDA-Messdaten für jede einzelne Person normalisiert, um die Daten interindividuell vergleichbar zu machen. Genutzt wird dafür folgende Berechnung (Boucsein, 2012, S. 181):

$$x = \text{Aktueller EDA_Wert} - \frac{\text{Mittelwert der EDA_Werte für diese Person}}{\text{Standardabweichung der EDA_Werte für diese Person}}$$

Im nächsten Schritt werden die ermittelten Werte x so transformiert, dass alle neuen Werte positiv sind und zudem ebenfalls im Wertebereich 0 bis 100 liegen. Genutzt wurde die folgende Formel (Boucsein, 2012, S. 181):

$$y = 50 + x \cdot 10$$

Durch die beschriebenen zwei Transformationsschritte der EDA-Messwerte wurde die Möglichkeit geschaffen, parallel zum Kennwert "Mimik" durch das Zuweisen von Vorzeichen einen zweiten Kennwert "Physio" zu bilden, der die Valenz aus der Mimikanalyse und das Arousal aus der EDA-Messung integriert.

Beide Kennwerte liegen in einem Wertebereich zwischen -100 und 100 und lassen sich auf der Grundlage der Daten aus Mimikanalyse und EDA-Messungen ermitteln. Eine positive bzw. negative Zahl ist gleichbedeutend mit einer positiven bzw. negativen emotionalen Valenz im Hinblick auf die Bewertung der Gebrauchstauglichkeit zu jedem Zeitpunkt der Messungen. Der Betrag des Wertes stellt jeweils die Beschreibung der emotionalen Dimension Arousal dar.

Forschungsfragen

Die beschriebenen Kennwerte liefern quantitative Daten, die auf objektiven Messungen im Labor beruhen. Die Aussagekraft dieser Daten hinsichtlich der Gebrauchstauglichkeit soll im Folgenden beurteilt werden. Dazu werden vier Forschungsfragen bearbeitet:

- Lassen sich gebrauchstaugliche von nicht-gebrauchstauglichen Gestaltungen einer Benutzungsoberfläche anhand der Kennwerte unterscheiden?
- Lassen sich in Verbindung mit Eye-Tracking einzelne Interaktionselemente mit besonders schlechter Gebrauchstauglichkeit identifizieren?

- Bringt die Integration der physiologischen Messwerte in den Kennwert "Physio" im Vergleich zum ausschließlich auf Daten zur Mimikanalyse beruhenden Kennwert "Mimik" einen Mehrwert?

Die drei Fragestellungen werden im Rahmen eines ersten Laborexperimentes mit einem Vergleich zwischen zwei Versionen einer einfachen Benutzungsoberfläche in Form eines Formulars untersucht. Die Ergebnisse werden dann in einem zweiten Experiment am Beispiel einer komplexen, ubiquitären Software auf Ihre Anwendbarkeit hin untersucht.

Methoden

Der Ablauf der beiden Versuche war so gestaltet, dass die Probanden gleich zu Beginn an der linken Hand mit zwei Elektroden für die Messung der EDA versorgt wurden, um bei Versuchsstart ein stabiles Mikroklima zwischen Elektrode und Haut zu gewährleisten und somit verfälschende Effekte auf die EDA-Messung zu vermeiden (Boucsein, 2012). Die Untersuchungen fanden unter stabilen Umweltbedingungen statt. Nun wurde den Probanden der Versuchsaufbau erklärt, mit dem die Nutzung einer interaktiven Software untersucht werden sollte. Die Probanden saßen an einem Computerarbeitsplatz, an dem sich auch das Gerät für berührungsloses Eye-Tracking und eine Videokamera für die Mimikanalyse befanden. Eine Einverständniserklärung für den geplanten Umgang mit den erhobenen Daten wurde unterschrieben. Nun wurden den Probanden die entsprechenden Aufgaben zur Interaktion mit der Software gestellt. Im Anschluss an die Erledigung jeder einzelnen Aufgabe wurden die Probanden gebeten, den Fragebogen "System Usability Scale SUS" (Brooke, 1996) für die durchgeführte Aufgabe auszufüllen.

Für die Datenauswertung wurden alle Datenquellen synchronisiert und anschließend in einer Datei mit einer Zeitreihe (10 Hz) zusammengeführt. Für jeden Zeitpunkt sind in der Datei folgende Sensordaten aufgeführt:

- Physiologische Messung: EDA-Leitwert in μS
- Videobasierte Mimikanalyse: vorherrschende Emotion zusammen mit dem zugehörigen, von der Software ermittelte Wert für die Stärke der Emotion, der aus den Daten zum Gesichtsausdruck errechnet wurde und Werte zwischen 0 und 100 aufweist
- Eye-Tracking: Bereich der graphischen Benutzungsoberfläche, auf den der Blick aktuell gerichtet war

Die ermittelten Daten wurden zur weiteren Auswertung in einer Datei gespeichert. Nun wurden die Kennwerte "Physio" und "Mimik" für jeden Zeitpunkt bestimmt.

Laborexperiment 1

Für dieses Experiment wurden zwei Versionen eines Formulars entwickelt, das folgende Informationen abfragt: Nachname, Vorname, Geburtsdatum, Geschlecht und aktueller Wohnort. Die erste Version sollte sich durch eine gute, die zweite durch eine schlechte Gebrauchstauglichkeit auszeichnen. Als Kriterium für die Erfüllung dieses Vorhabens wurde die positive Version in Übereinstimmung mit den Dialogprinzipien nach ISO 9241-110 gestaltet, während die negative Version an ausgewählten Stellen bewusst entgegen der Empfehlungen der Dialogkriterien gestaltet wurde:

- Angabe des Geburtsdatums: In der positiven Version (s. Abbildung 1 oben, erste Zeile) wurde das erforderliche Format der Datumseingabe dem Nutzer unmittelbar angezeigt (Dialogprinzip Selbstbeschreibungsfähigkeit erfüllt) und das Format TT.MM.JJJJ entspricht den kulturellen Präferenzen der Probanden (Dialogprinzip Erwartungskonformität erfüllt). Bei der negativen Version (s. Abbildung 1 unten, erste Zeile) werden keine Informationen zu dem erforderlichen Format angezeigt (Dialogprinzip Selbstbeschreibungsfähigkeit nicht erfüllt) und das erforderliche Format entspricht der nordamerikanischen und damit nicht der in Deutschland üblichen Form (Dialogprinzip Erwartungskonformität nicht erfüllt). Zudem bieten die Fehlermeldungen bei der Eingabe des Datums dem Nutzer keine Hilfe für eine korrekte Eingabe (Dialogprinzip Fehlertoleranz nicht erfüllt).
- Angabe des aktuellen Wohnorts: In der positiven Version (s. Abbildung 1 oben, zweite Zeile) ist in der angebotenen Dropdown-Liste eine Default-Voreinstellung "Deutschland" gewählt, die auf alle Probanden zutrifft (Dialogprinzip Aufgabenangemessenheit erfüllt). Bei der negativen Version (s. Abbildung 1 unten, zweite Zeile) ist das sichtbare Feld der Liste leer, und es werden keine weiteren handlungsleitenden Informationen wie beispielsweise "Bitte Land auswählen..." gegeben (Dialogprinzip Selbstbeschreibungsfähigkeit nicht erfüllt). Zudem ist die erscheinende Dropdown-Liste mit über 100 Ländern unsortiert, so dass ein langes Suchen unvermeidlich wurde (Dialogprinzip Aufgabenangemessenheit nicht erfüllt). Alle Probanden wohnten zum Zeitpunkt der Untersuchung in Deutschland. Die Angabe "Germany" war in der nicht alphabetisch geordneten Liste im unteren Drittel zu finden (Dialogprinzip Erwartungskonformität nicht erfüllt).

The image shows two versions of a form. The top version, labeled 'positive', has a 'Geburtsdatum' section with three dropdown menus for 'Tag', 'Monat', and 'Jahr', and an 'Aktueller Wohnort' section with a dropdown menu showing 'Deutschland'. The bottom version, labeled 'negative', has the same labels but uses empty text input fields for 'Geburtsdatum' and 'Aktueller Wohnort'.

Abb.1: Eingabefelder für Geburtsdatum und aktuellen Wohnort für die im Hinblick auf die Erfüllung der Dialogprinzipien positive (oben) und negative (unten) Version.

An dieser Untersuchung im Usability-Labor nahmen 22 Personen teil, die jeweils beide Formularversionen in ausbalancierter Reihenfolge nutzten. Die Datenerhebung über Eye-Tracking und videobasierte Emotionserfassung war mit teilweise erheblichen Datenlücken verbunden. Aus diesem Grund wurden die folgenden Analysen und die Ermittlung der Kennwerte lediglich für 10 Personen (5 Männer, 5 Frauen) im Alter zwischen 25 und 60 Jahren durchgeführt.

Die Tabelle 1 zeigt für jeden Probanden die Mittelwerte der während der Bearbeitung der Formulare ermittelten Kennwerte. Zudem sind die für die jeweilige Aufgabenerledigung über einen Fragebogen erfassten SUS-Werte aufgeführt.

Tab.1: Mittelwerte (MW) für die ermittelten Kennwerte und SUS-Werte

	Kennwert "Physio"		Kennwert "Mimik"		SUS-Wert	
	MW positive Version	MW negative Version	MW positive Version	MW negative Version	SUS negative Version	SUS positive Version
P01	50,1	45,9	87,2	81,3	92,5	20,0
P02	49,7	-1,6	65,5	1,7	85,0	30,0
P03	11,1	4,6	12,3	11,3	90,0	62,5
P04	10,0	1,3	4,0	-11,8	62,5	55,0
P05	37,3	-10,5	37,4	-13,1	77,5	12,5
P06	50,0	47,8	96,2	89,4	52,5	2,5
P07	6,5	2,1	11,9	4,0	87,5	30,0
P08	49,7	5,0	52,3	10,5	77,5	75,0
P09	43,0	22,7	53,1	23,0	65,0	47,5
P10	49,2	49,8	87,0	80,0	97,5	27,5
alle	35,7	16,7	50,7	27,6	78,8	36,3

Für alle sechs Datenreihen wurde ein Kolmogorov-Smirnov-Test durchgeführt, der die Bestätigung der Normalverteilung für alle Daten ergab. Nun wurden über alle Personen für den Kennwert "Physio", den Kennwert "Mimik" und den SUS-Wert jeweils ein einseitiger t-Test für abhängige Stichproben (Konfidenzintervall 95%) für die Mittelwertdifferenzen zwischen der positiven und der negativen Version durchgeführt. Alle drei Mittelwertdifferenzen sind signifikant. Aus den signifikanten Unterschieden zwischen den SUS-Werten der positiven und der negativen Version ergibt sich eine Bestätigung der auf der Gestaltung beruhenden Annahme, dass die positive Version tatsächlich eine bessere Gebrauchstauglichkeit aufweist als die negative. Bezogen auf die erste Forschungsfrage bedeuten die signifikanten Mittelwertdifferenzen für die Kennwerte "Physio" und "Mimik" eine Bestätigung der Hypothese, dass die ermittelten Kennwerte Interaktionen mit nachgewiesen unterschiedlicher Güte der Gebrauchstauglichkeit voneinander unterscheiden.

Zur Überprüfung der zweiten Forschungsfrage wurden die 10 niedrigsten Kennwerte über alle Probanden und alle durchgeführten Versuche bestimmt. Für diesen Vorgang wurde die Datenrate durch Mittelwertbildung von 10 Hz auf 1 Hz reduziert, um die ermittelte Situation einer wahrnehmbaren Interaktion zuordnen zu können und um eine Häufung von Extremwerten mit Bezug zu nur einem Interaktionsvorgang zu vermeiden. Über das Eye-Tracking konnte nun festgestellt werden, auf welche Interaktionselemente zu diesen Zeitpunkten der Blick der Probanden gerichtet war. Bis auf eine Ausnahme für den Kennwert "Physio" und eine Ausnahme für den Kennwert "Mimik" wurden zu den ermittelten Zeitpunkten die Interaktionselemente "Geburtsdatum" und "aktueller Wohnort" der negativen Version betrachtet. Dieses Ergebnis liefert ein Beispiel dafür, dass die vorgeschlagene Analyse der 10 niedrigsten Kennwerte in der ermittelten Zeitreihe mit hoher Wahrscheinlichkeit Hinweise auf tatsächliche Probleme mit der Gebrauchstauglichkeit liefert.

Mit Bezug auf die dritte Fragestellung kann festgehalten werden, dass der Kennwert "Physio" im Vergleich zum Kennwert "Mimik" keine bessere Unterscheidung bezüglich der Gebrauchstauglichkeit leistet. Für die Integration der physiologischen Daten konnte also kein Mehrwert gezeigt werden.

Laborexperiment 2

Die im Rahmen des Forschungsprojektes VENUS entwickelte, ubiquitäre Software MyGroup wurde in einer interdisziplinären Systemevaluation auf eine sozialverträgliche Gestaltung hin mit 22 potentiellen Nutzern untersucht. Wegen Datenausfall konnte die Auswertung lediglich für 12 Personen (9 Männer, 3 Frauen) im Alter zwischen 25 und 45 Jahren durchgeführt werden. Die Systemevaluation und MyGroup sind im einführenden Beitrag beschrieben.

Die im Laborexperiment 1 gewonnenen Erkenntnisse wurden für die Auswertung der im Rahmen der Systemevaluation im Usability Labor erhobenen Daten genutzt, um Interaktionselemente mit einer vergleichsweise niedrigen Qualität bezüglich der

Gebrauchstauglichkeit zu identifizieren. Dazu wurden mit Hilfe der Eye-Tracking Daten die 10 niedrigsten Werte des Kennwertes "Mimik" einzelnen Interaktionselementen zugeordnet. Allerdings war es nicht möglich, die 10 Situationen einfach als Probleme bei der Gebrauchstauglichkeit zu übernehmen. Vielmehr mussten die ermittelten Situationen einer kritischen Prüfung durch die Evaluierenden unterzogen werden. Folgende Punkte konnten im Rahmen der Auswertung als für die Gebrauchstauglichkeit kritische Interaktionselemente identifiziert werden:

- Das Diagramm zur Anzeige der Gesprächshistorie wurde nicht verstanden und Infoboxen auf dem Diagramm wurden für klickbare Elemente gehalten, was zu Irritationen führte.
- Eine Suche nach Kompetenzen anderer Teilnehmer an MyGroup wurde mehrfach auf einer Unterseite vermutet, die diese Art der Suche nicht ermöglicht.
- Die Ergebnisse der Suche nach Kompetenzen wurden nicht eindeutig als Ergebnisse der Suche erkannt.
- In zwei Fällen gab es einen Funktionsausfall bei der Angabe eines Gesprächswunsches.

Ein Nachweis anhand subjektiver Befragungsdaten der Probanden, dass es sich bei den genannten Punkten tatsächlich um Probleme mit der Gebrauchstauglichkeit handelt, konnte nicht geführt werden, weil die erhobenen SUS-Daten nicht fein genug nach einzelnen Interaktionselementen differenzieren.

Zusammenfassend lässt sich festhalten, dass die dargestellte Analyse des Kennwertes "Mimik" sich im Rahmen der Systemevaluation als hilfreicher Hinweisgeber für Probleme mit der Gebrauchstauglichkeit erwiesen hat.

Fazit und Ausblick

Der vorliegende Beitrag beschreibt einen Ansatz zur Messung von emotionaler Valenz und Arousal im Usability Labor sowie zur Bildung von Kennwerten für die objektive Beurteilung der Gebrauchstauglichkeit. Zudem werden über Eye-Tracking einzelne Interaktionselemente mit relativ schlechter Gebrauchstauglichkeit aufgefunden, um bei Verbesserungen gezielt ansetzen zu können. Es konnten starke Hinweise dafür gefunden werden, dass die beschriebene Kennwertbildung die Gebrauchstauglichkeit für ganze Interaktionsaufgaben und auch für einzelne Interaktionselemente abbilden. Die Ergänzung der Daten zu Valenz und Arousal aus der Mimikanalyse (Kennwert "Mimik") um Daten zum Arousal aus EDA-Messungen (Kennwert "Physio") erbrachte keinen messbaren Mehrwert. Dies mag damit zusammenhängen, dass Individuen physiologisch unterschiedlich auf Reize reagieren. Typabhängig können sich Erregungszustände beispielsweise in erster Linie mit einer Änderung der EDA oder über eine Änderung der Maße für die Herz-Rate verbunden sein (Lacey & Lacey, 1958). Neue eigene Forschungen beziehen sich auf eine typabhängige Kennwertbildung über EDA oder die Herzraten-

variabilität. Die Bildung von Korrelationskoeffizienten zu den Befragungsdaten in Untersuchungen mit mehr Probanden könnte weitere Erkenntnisse bringen, um die Integration von physiologischen Messdaten in die Kennwertbildung zu verbessern. Für die Auswertung der EDA-Daten sind in der vorliegenden Untersuchung tonische Maße zur Anwendung gekommen (Boucsein, 2012). Die Analyse der phasischen Anteile der EDA-Daten über eine Identifikation von lokalen Maxima (Benedek & Kaernbach, 2010) stellt ebenfalls einen interessanten Ansatz zur Datenauswertung dar. Insgesamt gesehen besteht eine zentrale Herausforderung darin, die ausgeprägten interindividuellen Unterschiede bei den messbaren körperlichen Reaktionen weiter zu erforschen, um eine typabhängige und gezielte Datenauswertung zu ermöglichen.

Die erheblichen Zeitaufwände für die Integration der Daten sprechen dafür, zukünftig einen effizienten Ansatz zu verfolgen, der so wenig Datenquellen wie möglich für die bestmöglichen Ergebnisse integriert. Auch in diesem Zusammenhang ist interessant, dass die Integration der physiologischen Daten keinen Mehrwert gebracht hat. Forschungen bezüglich der vielfältigen Möglichkeiten für physiologische Messungen im Usability-Labor zur Bewertung der Gebrauchstauglichkeit sollten zudem auch die Minimierung des Aufwandes für die Datenauswertung berücksichtigen.

Literatur

- Benedek, M. & Kaernbach, C. (2010). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, 47, 647-658.
- Boucsein, W. (2012). *Electrodermal activity*. New York: Springer.
- Brooke, J. (1996). SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.). *Usability Evaluation in Industry*. London: Taylor and Francis, 189-194.
- Ekman, P. (1970). Universal facial expressions of emotion. *California Mental Health Research Digest*, 8, 151-158.
- Lacey, J. I., & Lacey, B. C. (1958). Verification and extension of the principle of autonomic response-stereotypy. *The American Journal of Psychology*, 71(1), 50-73.
- Schmidt, L. (2009). Blickbewegungsanalyse zur software-ergonomischen Evaluation. In: *i-com - Zeitschrift für interaktive und kooperative Medien* 8,(1), S. 75-80
- Schmidt-Atzert, L. (1996). *Lehrbuch der Emotionspsychologie*. Stuttgart: Kohlhammer.
- Schmidtke, H. (2002). Vom Sinn und Unsinn der Messung psychischer Belastung und Beanspruchung. *Z. ARB. WISS.*, (56), 4-9.

Vertrauenswürdige Gestaltung von ubiquitären Systemen – Potentiale des Einsatzes von Methoden des NeuroIS

Matthias Söllner¹, Kay Behrenbruch², Holger Hoffmann¹ und Jan Marco Leimeister¹

Schlüsselwörter: Vertrauen, ubiquitäre Systeme, NeuroIS

Zusammenfassung

In den letzten Jahrzehnten war es in empirischen Studien disziplinübergreifend akzeptiert, sich fast ausschließlich auf Daten zu verlassen, die subjektive Einschätzung der Probanden widerspiegeln – z.B. Erheben durch Fragebögen oder Interviews. Die Einschränkung, dass subjektive Meinungsäußerungen von Probanden nicht zwangsweise identisch mit dem tatsächlichen Verhalten oder den wirklichen Emotionen sind, wurde weitgehend akzeptiert. In den letzten Jahren war jedoch das Aufkommen einer Vielzahl neuer methodischer Möglichkeiten zu beobachten, die es ermöglichen, die subjektiven Einschätzungen der Probanden durch objektive Daten anzureichern oder zu ersetzen. In der Wirtschaftsinformatik wird eines der zentralen Teilgebiete, das sich mit diesen Methoden befasst, als NeuroIS bezeichnet. Ziel dieses Beitrags ist es am Beispiel des ubiquitären Systems MyGroup darzustellen, welche Potentiale der Einsatz von Methoden des NeuroIS für die vertrauenswürdige Gestaltung von ubiquitären Systemen bietet. Konkret fokussiert sich dieser Beitrag auf das Potential des Einsatzes von Eye-Tracking, videobasierter Emotionserkennung und physiologischen Messungen zur Beurteilung der Wirksamkeit existierender und Identifikation neuer Designkomponenten.

Einleitung

Zahlreiche Forschungsarbeiten haben die Bedeutung von Vertrauen in verschiedenen Forschungsgebieten der Wirtschaftsinformatik gezeigt. Einige Beispiele sind E-Commerce (Gefen & Straub, 2004), die Adoption neuer Technologien (Gefen et al., 2003) und die virtuelle Zusammenarbeit (Jarvenpaa et al., 2004). Als besonders wichtig betonen viele Forscher die Bedeutung des Vertrauensaufbaus (Gefen et al., 2003; Resatsch et al., 2008), der Vertrauensunterstützung (Leimeister et al., 2005) und der Identifikation von Faktoren, die Vertrauen erzeugen (Bart et al., 2005). Resultat dieser Wichtigkeit waren einige Arbeiten, die der Frage nachgehen, wie verhaltensorientierte Erkenntnisse zu Vertrauen dazu verwendet werden können,

¹ Universität Kassel, Fachgebiet Wirtschaftsinformatik, Pfannkuchstr. 1, 34121 Kassel, {Leimeister, Holger.Hoffmann, Soellner}@uni-kassel.de

² Universität Kassel, Fachgebiet Mensch-Maschine-Systemtechnik, Mönchebergstr. 7, 34125 Kassel, K.Behrenbruch@uni-kassel.de

vertrauenswürdiger Informationssysteme zu gestalten (Patrick et al., 2005; Söllner & Leimeister, 2011). Eine Limitation dieser Arbeiten ist es jedoch, dass die Beurteilung der gestiegenen Vertrauenswürdigkeit nicht auf Basis objektiver Messgrößen erfolgt – zum Beispiel einer höheren objektiven Nutzung des Systems oder objektiv erfassten emotionalen Reaktionen auf die Variationen im Design. Stattdessen werden Laborexperimente verwendet, in denen verschiedene Varianten von System miteinander verglichen werden. Die Daten für den Vergleich werden meist mit Hilfe eines Fragebogens erhoben, mit Hilfe dessen die Teilnehmer ex post ihre Eindrücke dokumentieren sollen (Söllner & Leimeister, 2011).

Auch wenn dieses Vorgehen auch aktuell noch als State-of-the-Art bezeichnet werden kann, so gibt es doch neue methodische Entwicklungen, die Schwächen der bisherigen Evaluationen beseitigen können (Dimoka et al., 2011). So ist es zum Beispiel möglich objektiv zu erfassen, welche GUI-Elemente von Nutzern wie intensiv angesehen werden. So kann überprüft werden, ob eine Veränderung, die vorgenommen wurde, um das Vertrauen des Nutzers zu erhöhen, überhaupt wahrgenommen wird. Des Weiteren ist es möglich, mit Hilfe von Emotionserkennungssoftware in Echtzeit die emotionalen Reaktionen der Nutzer bei der Verwendung des Systems zu erfassen. Die Triangulation beider Daten ermöglicht es dann noch, die Emotionen konkreten GUI-Elementen zuzuweisen.

Ziel dieses Beitrags ist es, am Beispiel des ubiquitären Systems MyGroup zu zeigen, wie vergleichbare Methoden eingesetzt werden können, um so genannten vertrauensunterstützende Komponenten zu bewerten und Situationen zu identifizieren, in denen die Implementierung einer solchen Komponente hilfreich wäre.

Um dieses Ziel zu erreichen, ist der Beitrag wie folgt gegliedert. Zuerst wird das ubiquitäre System MyGroup vorgestellt. Anschließend wird kurz auf die theoretischen Grundlagen zu Vertrauen eingegangen und vertrauensunterstützende Komponenten für MyGroup abgeleitet. Danach werden die Grundlagen des NeuroIS eingeleitet und die in diesem Beitrag verwendeten Methoden näher beschrieben. Als nächstes wird die Forschungsmethodik dargelegt, bevor die Ergebnisse präsentiert werden. Der Beitrag schließt mit einem Fazit und zukünftigem Forschungsbedarf.

MyGroup

Die im Rahmen des Forschungsprojektes VENUS entwickelte, ubiquitäre Software MyGroup wurde in einer interdisziplinären Systemevaluation auf eine sozialverträgliche Gestaltung hin mit 24 potentiellen Nutzern untersucht. Die Systemevaluation und MyGroup sind in dem vorangegangenen Beitrag "Systemevaluation" beschrieben.

Vertrauensunterstützende Komponenten

Die Bedeutung von Vertrauen im Kontext der Adoption von Informationssystemen wurde mehrfach in unterschiedlichen Studien herausgestellt (siehe Söllner und Leimeister (forthcoming) für einen Überblick über die verschiedenen empirisch quantitativen Arbeiten, die zwischen 1995 und 2012 in hochrangigen Zeitschriften der Wirtschaftsinformatik veröffentlicht wurden). Der Hauptfokus der Forschung liegt weiterhin auf der Generierung verhaltensorientierter Forschungserkenntnisse, z.B. zur Bedeutung und Zusammensetzung von Vertrauen in unterschiedlichen Kontexten (siehe z.B. Gefen et al. (2003), Gefen & Pavlou (2011) und Söllner et al. (2012)). Nichtsdestotrotz existieren auch Artikel, die thematisieren, wie die Vielzahl an verhaltensorientierten Erkenntnisse verwendet werden kann, um vertrauenswürdiger Informationssysteme zu entwickeln. Der Beitrag von Söllner et al. (2011) fokussiert hier explizit ubiquitäre Informationssysteme und dient daher als Grundlage für den vorliegenden Beitrag und wurde auch für die Ableitung von vertrauensunterstützenden Komponenten für MyGroup verwendet. Tabelle 1 fasst die für diesen Beitrag wichtigen Ergebnisse der Anwendung der in Söllner et al. (2011) dargestellten Methode im Kontext von MyGroup zusammen und präsentiert auch zwei vertrauensunterstützende Komponenten, die in MyGroup implementiert wurden. Auf diesen beiden Komponenten liegt auch das Hauptaugenmerk der späteren Evaluation.

Tab.1: Vertrauensunterstützende Komponenten in MyGroup nach Söllner et al. (2011).

	Beschreibung	Vertrauens- dimension	Vertrauens- determinante
Vertrauens- unterstützende Komponente 1	Informationen, die verdeutlichen, warum verschiedene Profile in sozialen Netzwerken mit MyGroup verbunden werden sollten.	Prozessqualität	Verständlichkeit
Vertrauens- unterstützende Komponente 2	Einstellungsmöglichkeiten, welche der eigenen Profilinformationen wie detailliert von verschiedenen Gruppen von anderen Nutzern eingesehen werden können.	Prozessqualität	Kontrolle

Die erste Komponente ist in MyGroup in Form von kontextsensitiv angezeigten Informationen realisiert: Wenn der Mauszeiger in die Nähe des entsprechenden Eingabefeldes bewegt wird, erscheint eine Infobox, die verdeutlichte, warum die verschiedenen Profile in sozialen Netzwerken mit MyGroup verbunden werden sollten (Abbildung 1). Die zweite vertrauensunterstützende Komponente ist in

MyGroup als Auswahlliste realisiert, die die Sichtbarkeit verschiedener privater Informationen in fünf Abstufungen einstellbar macht (Abbildung 2). Zudem werden bei Mausbewegungen zu den Auswahlpunkten Zusatzinformationen angezeigt (Abbildung 2, rechte Seite).

Abb. 1: Vertrauensunterstützende Komponente 1

Abb. 2: Vertrauensunterstützende Komponente 2

NeuroIS

NeuroIS ist ein Forschungsfeld in der Wirtschaftsinformatik, das darauf basiert, Erkenntnisse auf dem Gebiet der kognitiven Neurowissenschaft einzusetzen, um zentrale Forschungsfelder der Wirtschaftsinformatik effektiver beforschen zu können (Dimoka et al., 2011). Hierbei kommt eine Vielzahl von unterschiedlichen Methoden und Tools zum Einsatz (für einen Überblick siehe Dimoka et al. (2012)). Im vorliegenden Beitrag liegt das Erkenntnisinteresse darauf, die implementierten vertrauensunterstützenden Komponenten zu bewerten sowie Situationen zu identifizieren, in denen weitere Komponenten hilfreich zu sein scheinen. Um dieses Ziel zu erreichen, werden verschiedene Methoden eingesetzt: Eye-Tracking, videobasierte Emotionserkennung, physiologische Messungen und Verhaltensbeobachtung.

Mit Hilfe von Eye-Tracking soll analysiert werden, (1) ob die vertrauensunterstützenden Komponenten von den Nutzern wahrgenommen werden und (2) wie intensiv sie wahrgenommen werden, also z.B. wie lange ein Nutzer die Komponenten verwendet (z.B. wie intensiv die Informationstexte gelesen werden).

Mit Hilfe aller Daten soll analysiert werden, (1) welche Emotionen die Nutzer zeigen, wenn Sie eine vertrauensunterstützende Komponente wahrnehmen bzw. wenn Sie diese verwenden und (2) ob sie an anderen Stellen im Interaktionsprozess negative Emotionen zeigen, die mit Hilfe von zusätzlichen vertrauensunterstützten Komponenten verhindert werden könnten.

Forschungsmethodik

Die Laboruntersuchungen wurden in einem voll ausgerüsteten Usability Labor durchgeführt. Die Probanden wurden an der linken Hand mit Elektroden zur Messung der elektrodermalen Aktivität (EDA) ausgestattet. Die Probanden saßen an einem Schreibtisch und interagierten während der Aufgabenerledigung mit der Software MyGroup über einen Flachbildschirm, Maus und Tastatur. Während der Versuchsdauer wurde ein Video von den Bildschirminhalten mitgeschnitten. Die mit einem berührungslosen Eye-Tracker aufgenommenen Daten zu Blickpfaden und Fixationen wurden gespeichert und in dem Video von der Bildschirminhalten angezeigt. Zudem wurden während der Interaktion Videoaufnahmen vom Gesicht der Probanden aufgenommen und sowohl zur Verhaltensbeobachtung im Zusammenhang mit Audiodaten als auch zur videobasierten Emotionserkennung verwendet.

An den Versuchen nahmen 22 der 24 Teilnehmer der Systemevaluation teil. Wegen technischer Probleme konnten allerdings lediglich 12 Datensätze von 9 Männer und 3 Frauen im Alter zwischen 25 und 45 Jahren für die Auswertung herangezogen werden. Alle erhobenen Daten wurden synchronisiert gespeichert.

Den Probanden wurden zwei Aufgaben für die Interaktion mit MyGroup gestellt:

- Bitte pflegen Sie in dem Test-Account die Account Settings für „Social Accounts“.
- Bitte pflegen Sie in dem Test-Account die Account Settings für „Privacy Settings“.

Die beiden Aufgaben wurden nacheinander bearbeitet. Die Probanden wurden zudem gebeten, beim Erledigen der Aufgaben so vorzugehen wie sie es bei Ihrem eigenen MyGroup-Account tun würden. Die Probanden signalisierten das Ende der Aufgabe durch das Zurückkehren auf den Startbildschirm von MyGroup.

Für einen späteren Abgleich der objektiv erfassten Daten zur EDA, zur Blickzuwendung, zu den in der Mimik erkannten Emotionen und zur Verhaltensbeobachtung mit subjektiven Einschätzungen der Probanden wurde nach der Beendigung der Aufgaben ein Fragebogen ausgefüllt. Der Fragebogen enthielt folgende Fragen, die auf einer 5-stufigen Likert Skala (1 = trifft gar nicht zu ... 5 = trifft voll und ganz zu) beantwortet wurden:

- Die angezeigten Erläuterungen für die in den Account Settings erfragten Informationen sind mir wichtig.

- Die aktuelle Umsetzung der Anzeige der Erläuterungen ist gelungen.
- Die Möglichkeit genau zu definieren welcher Nutzer welche Information über mich einsehen darf ist mir wichtig
- Die aktuelle Umsetzung der Möglichkeit zu bestimmen wer welche Information sehen darf ist gelungen.

Ergebnisse

Im Fokus der Datenauswertung standen zwei in MyGroup realisierte vertrauensunterstützende Komponenten (vgl. Tabelle 1, Abbildung 1 und Abbildung 2).

Durch die Analyse der Eye-Tracking-Daten konnte festgestellt werden, dass die erste vertrauensunterstützende Komponente in Form der kontextsensitiv erscheinenden Infoboxen zur Verwendung der optional anzugebenden Daten für die sozialen Netzwerke (Abbildung 1) von den Probanden durchschnittlich 5 Sekunden betrachtet wurde. Da sich diese 5 Sekunden im Ablauf der Aufgabenbearbeitung auf bis zu 7 Infoboxen verteilten, wurde vermutet, dass die angebotenen Texte nicht bei allen Blickzuwendungen auch gelesen wurden. Eine Überprüfung anhand der Bildschirmmitschnitte mit Blickpfaden bestätigte diese Vermutung: Nur 3 der 12 Probanden hatten die Infotexte tatsächlich gelesen. Für die übrigen Zeiten, in denen der Blick der Probanden auf die Infoboxen gerichtet war, müssen also andere Gründe als das Lesen der Texte angenommen werden. In diesem Zusammenhang liegt die Vermutung nahe, dass die Infoboxen Bildschirminhalte verdecken, die eigentlich angesehen werden sollten. Diese Vermutung liegt nahe, weil in der getesteten Realisierung von MyGroup die Infoboxen vor anderen Eingabefeldern erscheinen (Abbildung 1, linke Seite). Für die Überprüfung dieser Vermutung wurden die emotionsbezogenen Daten ausgewertet. Dazu wurde ein Kennwert genutzt, der die Wertigkeit und die Stärke der emotionalen Reaktionen bei der Nutzung von Software im Hinblick auf deren Gebrauchstauglichkeit in einem Wert (Wertebereich -100 bis 100) integriert. Dieser Kennwert wird in dem vorangehenden Beitrag "Gebrauchstauglichkeit im Labor objektiv evaluieren" ausführlich beschrieben und diskutiert. Tatsächlich wurde in den Zeiträumen, in denen die Infoboxen angeschaut wurden ein deutlich niedrigerer Kennwert (-31) festgestellt, als für die restliche Zeit der Interaktion (-3). Die emotionalen Reaktionen, wenn der Blick auf die Infoboxen gerichtet ist, sind also vermehrt negativ (z.B. ärgerlich). Im Verbund mit der Beobachtung, dass die angebotenen Texte nur selten gelesen werden, ergibt sich eine Bestätigung der Vermutung, dass die Infoboxen stören, weil sie Elemente verdecken, die eigentlich angeschaut werden sollen.

Die Auswahlmöglichkeiten als Realisierung der zweiten vertrauensunterstützenden Komponente (Tabelle 1 und Abbildung 2) nahmen fast den gesamten Bildschirm ein, und es wurden nur sehr selten andere Elemente als die Liste der Auswahlmöglichkeiten angesehen. Aus diesem Grund konnte das Eye-Tracking nicht genutzt werden, um die Betrachtung einzelner Elemente der vertrauensunterstützenden Komponente zu erkennen und zu unterscheiden. Nach einer Analyse der

emotionalen Reaktionen wurde festgestellt, dass besonders negative emotionale Reaktionen (niedriger Kennwert) in mehreren Fällen mit einer Suchbewegung des Blickes auf dem Bildschirm verbunden waren. Zudem konnte auch bei dieser Aufgabe festgestellt werden, dass die auf Mausbewegungen hin erscheinenden Infoboxen mit Zusatzinformationen zu den Auswahlmöglichkeiten selten gelesen wurden, andere Auswahlmöglichkeiten verdeckten und mit negativen emotionalen Reaktionen (niedriger Kennwert) verbunden waren. Die Suchbewegungen der Augen und die angesprochene Verdeckung verbunden mit den negativen emotionalen Ausprägungen wurden mit einer Unübersichtlichkeit der Auswahlmöglichkeiten in Zusammenhang gebracht. Diese Unübersichtlichkeit konnte an den sich bei jeder Auswahlmöglichkeit wiederholenden und gleichen Textbausteinen "Visible for" sowie an den verdeckenden Infoboxen festgemacht werden. Die Abbildung zwei zeigt lediglich die Auswahlmöglichkeiten für die Rubrik "Profile Settigs". Auf der gleichen Seite werden 5 weitere Kategorien mit den gleichen 5 Auswahlmöglichkeiten angeboten, was die Redundanz der Informationen und die Unübersichtlichkeit erhöht.

Die Auswertung der subjektiven Fragebogendaten für die 12 Probanden ergab im Zusammenhang mit den dargestellten objektiven Daten interessante Ergebnisse. Bezogen auf die erste vertrauensunterstützende Komponente gaben die Probanden an, diese wichtig zu finden (Median=4 auf der Skala von 1 bis 5). Die objektive Messung ergab allerdings, dass die Informationstexte kaum gelesen wurden. Hier zeigt sich wohl eine Schwäche bei der Konzeption der Untersuchungen: Da die Probanden MyGroup bereits kannten und nutzten als sie den Test durchführten, hatte sie die für sie eigentlich wichtigen Informationen wohl schon zu einem früheren Zeitpunkt gelesen, im Experiment erinnert und nicht noch einmal gelesen. Auch die zweite vertrauensunterstützende Komponente wurde von den Probanden als wichtig eingestuft (Median=5). Zu dieser Aussage findet sich in den Ergebnissen der Auswertung der objektiven Labordaten weder eine Bestätigung noch ein Widerspruch. Ebenfalls erfragt wurde die Einschätzung der Qualität der Umsetzung der beiden vertrauensunterstützenden Komponenten. Sowohl für die erste (Median=1) als auch für die zweite (Median=2) vertrauensunterstützende Komponente ergaben sich negativen Bewertungen der Qualität der Umsetzung. Dieses Ergebnis korrespondiert mit den ermittelten Problemen der Verdeckung und der Unübersichtlichkeit.

Bezogen auf die Identifikation von Interaktionselementen, bei denen zusätzliche vertrauensunterstützende Komponenten förderlich sein könnten, wurde ein erweiterter Datensatz, der sich auch auf andere Interaktionsmöglichkeiten mit MyGroup bezieht, auf Situationen untersucht, in denen eine besonders hohe Ausprägung der beiden emotionalen Ausprägungen "ärgerlich" und "überrascht" gemessen wurde. Die Annahme war, dass eine gleichzeitige hohe Ausprägung bei diesen beiden Emotionen auf Situationen hinweisen könnte, die mit einem niedrigen Vertrauensniveau verbunden sind. Für diese intuitive Annahme existieren keine uns bekann-

ten Vorerkenntnisse. Die Analyse ergab im Zusammenhang mit dem Eye-Tracking und den Video- und Audioaufzeichnungen der Probanden interessante Situationen bei zwei einzelnen Probanden, die sich allerdings bei anderen Probanden nicht bestätigen ließen. Es handelte sich dabei um eine Irritation, weil Änderungen am Profil nicht bestätigt wurden und um fehlende Informationen zum Zeitpunkt der erfolgten Gespräche beim Betrachten eines Diagramms zur Darstellung der persönlichen Gesprächshistorie.

Fazit und zukünftiger Forschungsbedarf

In diesem Beitrag konnte am Beispiel der Evaluation einer ubiquitären Software gezeigt werden, dass die objektiven Methoden des NeuroIS geeignet sind, um in empirischen Studien auswertbare Daten zu gewinnen. Allerdings war es für eine sinnvolle Interpretation der Messdaten zu Physiologie, Blickbewegung und sich im Gesicht widerspiegelnden Emotionen auch notwendig, das Verhalten der Probanden zu beobachten und einen Abgleich mit subjektiven Befragungen durchzuführen. Zudem muss der technische Aufwand eines voll ausgestatteten Usability Labors im Zusammenhang mit den notwendigen Ressourcen zur Auswertung der komplexen Daten bei einer Beurteilung der Praxistauglichkeit des vorgestellten Vorgehens beachtet werden.

Die gemeinsame Auswertung von objektiven Daten aus den Methoden des NeuroIS und den subjektiven Fragebogendaten hat sich als geeignet dafür erwiesen zu unterscheiden, ob die Komponente an sich oder ob deren Umsetzung gut bzw. verbesserungswürdig sind. Man kann also Anforderung und Gestaltung sehr gut voneinander getrennt evaluieren.

Eine Herausforderung bei der Auswertung der objektiv erhobenen Daten liegt in der stark ausgeprägten, interindividuellen Variation der Messgrößen. Eine Typisierung der physiologischen und emotionalen Reaktionen bei der Interaktion mit Software stellt ein interessantes Forschungsfeld dar, das bei der Interpretation von objektiv erhobenen Labordaten am Menschen sehr hilfreich sein kann. Hierzu liegen beispielsweise für die Physiologie relativ alte Erkenntnisse vor (z.B. Lacey & Lacey, 1958), die im Zusammenhang NeuroIS aufgegriffen und ausgebaut werden sollten.

Literatur

- Bart, Y., Shankar, V., Sultan, F., & Urban, G. L. (2005). Are the Drivers and Role of Online Trust the Same for All Web Sites and Consumers? *Journal of Marketing*, 69(4), 133-152.
- Dimoka, A., Banker, R. D., Benbasat, I., Davis, F. D., Dennis, A. R., Gefen, D. & Weber, B. (2012). On the Use of Neurophysiological Tools in IS Research: Developing a Research Agenda for NeuroIS. *MIS Quarterly*, 36(3), 679-A619.

- Dimoka, A., Pavlou, P. A., & Davis, Fred D. (2011). NeuroIS: The Potential of Cognitive Neuroscience for Information Systems Research. *Information Systems Research*, 22(4), 687-702. doi: 10.1287/isre.1100.0284
- Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in Online Shopping: An Integrated Model. *MIS Quarterly*, 27(1), 51-90.
- Gefen, D., & Pavlou, Paul A. (2011). The Boundaries of Trust and Risk: The Quadratic Moderating Role of Institutional Structures. *Information Systems Research*.
- Gefen, David, & Straub, D. W. (2004). Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services. *Omega*, 32(6), 407-424.
- Jarvenpaa, S. L., Shaw, T. R., & Staples, D. S. (2004). Toward Contextualized Theories of Trust: The Role of Trust in Global Virtual Teams. *Information Systems Research*, 15(3), 250-264.
- Leimeister, J. M., Ebner, W., & Krcmar, H.. (2005). Design, Implementation, and Evaluation of Trust-Supporting Components in Virtual Communities for Patients. *Journal of Management Information Systems*, 21(4), 101-135.
- Lacey, J. I., & Lacey, B. C. (1958). Verification and extension of the principle of autonomic response-stereotypy. *The American journal of psychology*, 71(1), 50-73.
- Patrick, A. S., Briggs, P., & Marsh, S. (2005). Designing systems that people will trust. In L. Cranor & S. Garfinkel (Eds.), *Security and Usability: Designing Secure Systems That People Can Use*: O'Reilly.
- Resatsch, F., Sandner, U., Leimeister, J. M., & Krcmar, H. (2008). Do Point of Sale RFID-Based Information Services Make a Difference? Analyzing Consumer Perceptions for Designing Smart Product Information Services in Retail Business. *Electronic Markets*, 18(3), 692-709.
- Söllner, M., & Leimeister, J. M. (2011). Vertrauensunterstützung für ubiquitäre Systeme – von der Vertrauens Theorie zu Anforderungen an ubiquitäre Systeme. Paper presented at the 9. Berliner Werkstatt Mensch-Maschine-Systeme, Berlin, Germany.
- Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A., & Leimeister, J. M. (2012). Understanding the Formation of Trust in IT Artifacts. Paper presented at the Proceedings of the International Conference on Information Systems (ICIS) 2012, Orlando, Florida, USA.
- Söllner, M., & Leimeister, J. M. (forthcoming). What We Really Know About Antecedents of Trust. In D. Gefen (Ed.), *Psychology of Trust: New Research*: Nova Science Publishers.

Interdisziplinäre Sicherheitsanalyse

Olga Kieselmann¹, Nils Kopal¹ und Arno Wacker¹

Schlüsselwörter: Sicherheitsanalyse, Interdisziplinarität, Ubiquitäre Systeme, Simulationsstudie

Einführung

Im Rahmen der Evaluation einer ubiquitären Anwendung wird in der Regel auch die Erfüllung der Sicherheitsanforderungen durch Sicherheits-Spezialisten überprüft. In einem rein disziplinären Vorgehen wird dies vollständig unabhängig von den anderen Disziplinen (z.B. Benutzervertrauen oder rechtliche Anforderungen) durchgeführt. Die dabei entdeckten Sicherheitsmängel berücksichtigen häufig nicht die Endnutzer und können Auswirkung auf andere Disziplinen haben. Daher schlagen wir in diesem Beitrag eine neue interdisziplinäre Sicherheitsanalyse vor, welche den Endnutzer und die Abhängigkeiten zu anderen Disziplinen berücksichtigt. Dies erreichen wir, indem wir jeden Schritt der klassischen Sicherheitsanalyse interdisziplinär durchführen und erkannte Bedrohungen während einer Systemevaluation als konkrete Angriffe simulieren. Darüber hinaus zeigen wir die konkrete Anwendung unserer interdisziplinären Sicherheitsanalyse am Beispiel einer prototypischen Anwendung.

Motivation

Die IT-Sicherheitsanalyse stellt für ubiquitäre Systeme eine besondere Herausforderung dar, da in solchen Systemen in der Regel viele Sensoren und Aktuatoren beteiligt sind, die neuartige Sicherheitsprobleme mit sich bringen. In solchen Systemen werden Sensordaten häufig über drahtlose Netzwerke ausgetauscht, wodurch bei mangelhaften Sicherheitsmechanismen ein hohes Schadenspotential entsteht. So könnte ein potentieller Angreifer mit Hilfe von gesammelten Sensordaten aus dem Haus des Nutzers feststellen, wann er schläft bzw. wann er nicht zuhause ist, um einen geeigneten Zeitpunkt für einen Einbruch auszuwählen. Um solchen Missbrauch des Systems vorzubeugen, muss nach dem Fertigstellen des Systems eine Sicherheitsanalyse durchgeführt werden.

Die klassische Sicherheitsanalyse eines IT-Systems wird durch Spezialisten vor seinem realen Einsatz durchgeführt. Die Praxis zeigt, dass viele Schwachstellen erst entdeckt werden nachdem das System eingesetzt wird, da u. a. der Faktor Mensch nicht ausreichend berücksichtigt wurde. Zum Beispiel werden Bedrohungen, die von Benutzern ausgehen, wie z.B. Phishing, häufig beim Designprozess einer Software nicht betrachtet. Wir schlagen eine neue interdisziplinäre Sicher-

¹ Universität Kassel, Fachgebiet Angewandte Informationssicherheit, Pfannkuchstr. 1, 34121 Kassel, {Arno.Wacker, Kieselmann, Nils.Kopal}@uni-kassel.de

heitsanalyse vor, welche sich von der klassischen Sicherheitsanalyse insbesondere durch folgende zwei Punkte unterscheidet:

- Alle Schritte der Sicherheitsanalyse werden interdisziplinär durchgeführt, wodurch neben den technischen Aspekten auch Aspekte wie Benutzerfreundlichkeit oder das Vertrauen in die Anwendung berücksichtigt werden.
- Durch die Durchführung einer Simulationsstudie als Teil der Systemevaluation, welche den Einsatz der Anwendung unter realen Bedingungen simuliert, wird zusätzlich der Faktor Mensch berücksichtigt, wodurch auch Schwachstellen erkannt werden, die einen nicht-technischen Ursprung haben.

Im Folgenden stellen wir zunächst die klassische Sicherheitsanalyse vor, danach beschreiben wir unsere interdisziplinäre Lösung und zeigen einige Beispiele aus einer durchgeführten Simulationsstudie, auf Grundlage der ubiquitären Anwendung MyGroup. Den Schluss des Beitrags bildet eine kurze Zusammenfassung der erzielten Ergebnisse.

Klassische Sicherheitsanalyse

Die klassische Sicherheitsanalyse beginnt mit der Design-Analyse des implementierten Systems, um einen Überblick über den Systemaufbau sowie die einzelnen Komponenten des Systems zu bilden und somit die möglichen Schwachstellen zu entdecken. Dabei wird ein besonderes Augenmerk auf den Datenfluss innerhalb des Systems gelegt. Hierzu werden die Dokumentation sowie der Source-Code herangezogen. Das gesamte Verfahren der klassischen Sicherheitsanalyse kann in folgende Schritte eingeteilt werden (Paulus, 2011):

1. Analyse/Identifizierung der Sicherheitsziele: In diesem Schritt werden die zu schützende Informationen bzw. Komponenten identifiziert. Die Sicherheitsziele werden in der Regel in der Anforderungserhebung-Phase definiert. In der Evaluations-Phase dienen sie der Überprüfung der Umsetzung der festgelegten Sicherheitsziele in dem implementierten System.
2. Design und Code Review/Bestandsaufnahme: Nachdem im 1. Schritt die Sicherheitsziele ermittelt wurden, wird das System als Ganzes betrachtet. Es wird analysiert, wie das System aufgebaut ist, aus welchen Komponenten es besteht und wie die Schnittstellen zwischen den Komponenten realisiert sowie welche Sicherheitsmaßnahmen umgesetzt wurden. Die Analyse in dieser Phase wird anhand der vorliegenden Dokumentation durchgeführt, d. h. sämtliche Dokumentation sowie der Quellcode stehen zur Verfügung. Das Ziel dieser Phase ist es, zu prüfen ob die Sicherheitsziele umgesetzt wurden.
3. Erkennung von Schwachstellen: Mit den aus der Bestandsaufnahmen gewonnen Informationen über das System wird analysiert, welche Schwachstellen das System aufweist und wie diese für Angriffe ausgenutzt werden können. Dafür wird das System in Einzelteile zerlegt, die gesondert be-

trachtet werden müssen. Hierzu werden unterschiedliche Techniken eingesetzt wie z.B. Erstellung von Angriffsvektoren. Das Ziel dieser Phase ist es, zu prüfen ob, trotz der umgesetzten Schutzmaßnahmen, Sicherheitsprobleme im System bestehen, die für Angriffe ausgenutzt werden könnten. Da es selten möglich ist, ein System gegen jede mögliche Bedrohung zu schützen, wird zusätzlich eine Risikobewertung durchgeführt, um das System gegen die wichtigsten und kritischsten Bedrohungen zu schützen.

4. Sicherheitstests/Angriffssimulationen: Die in dem vorangegangenen Schritt erkannten Schwachstellen werden zur Simulation von Angriffen genutzt, um zu prüfen, ob die Angriffe praktisch durchführbar sind und welche Auswirkungen diese mit sich bringen würden. Das Ziel dieser Phase ist es, die Ausmaße der erkannten Schwachstellen bzw. Bedrohungen zu ermitteln sowie ggf. die in vorangegangenen Schritten unentdeckten Bedrohungen zu erkennen.
5. Schutzmaßnahmen: Im Rahmen dieses Schrittes werden anhand der Ergebnisse der Risikobewertung sowie der Sicherheitstests die möglichen Bedrohungen aufgezeigt sowie die entsprechenden Gegenmaßnahmen ausgearbeitet.

Das Ziel der Sicherheitsanalyse ist, zu prüfen, ob die Sicherheitsanforderungen umgesetzt wurden sowie ob die definierten Sicherheitsziele erreicht wurden. Zudem wird mittels des Sicherheitstests das System auf seine Widerstandsfähigkeit geprüft, um mögliche Restrisiken aufzudecken. Die Sicherheitsanalyse wird mit einem Vorschlag für mögliche Sicherheitsmaßnahmen gegen die entdeckten Schwachstellen abgeschlossen.

Interdisziplinäre Sicherheitsanalyse

Im Folgenden gehen wir auf die Unterschiede zwischen der klassischen Sicherheitsanalyse und unserer entwickelten interdisziplinären Lösung ein. Als grundlegendes Vorgehen haben wir die klassische Sicherheitsanalyse genutzt und diese um interdisziplinäre Aspekte erweitert. Diese werden im Folgenden in dem jeweiligen Schritt kurz aufgezeigt:

1. Analyse/Identifizierung der Sicherheitsziele: Neben den klassischen Sicherheitszielen, die durch die zu untersuchende Anwendung vorgegeben sind, werden bei der interdisziplinären Sicherheitsanalyse gleichzeitig auch Ziele anderer Disziplinen berücksichtigt. Rechtliche Vorgaben können z.B. zu stärkeren Sicherheitsanforderungen führen als es die Entwickler für die Anwendung vorgesehen hatten. Gleichzeitig wirken Anforderungen aus dem Bereich der Benutzerschnittstellen gegensätzlich, da eine höhere Sicherheit zumeist auf Kosten der Benutzbarkeit einhergeht. Ebenso können Anforderungen aus dem Bereich Vertrauen dazu führen, dass zusätzliche sicherheitsrelevant erscheinende Funktionen implementiert

werden müssen, obwohl sie die Sicherheit aus technischer Sicht nicht erhöhen.

2. Design und Code Review/Bestandsaufnahme: Wie bei einer klassischen Sicherheitsanalyse wird bei der interdisziplinären Sicherheitsanalyse eine Code-Analyse durchgeführt, um zu prüfen, ob die Sicherheitsziele umgesetzt wurden. In diesem Schritt ist die klassische und interdisziplinäre Sicherheitsanalyse identisch.
3. Erkennung von Schwachstellen: Genauso wie bei der klassischen Sicherheitsanalyse werden, basierend auf der Bestandsaufnahme aus dem vorherigen Schritt, die potentiellen Schwachstellen identifiziert. Ein ubiquitäres System besteht aus einer Vielzahl von vernetzten Komponenten. Daher wird darüber hinaus bei der interdisziplinären Sicherheitsanalyse angenommen, dass einzelne Komponenten unter der Kontrolle eines Angreifers stehen. Dies wird angenommen, ohne dass eine konkrete Schwachstelle bei den jeweiligen Komponenten festgestellt wurde, die zu solch einem Angriff geführt haben könnte. So kann z.B. angenommen werden, dass ein Server, auf dem eine Datenbank läuft, von dem Angreifer unter seine Kontrolle gebracht wird, wodurch er vollständigen Zugang zu allen Daten auf dem System hat. Ebenso kann ein öffentlich zugänglicher Sensor mit relativ kleinem Aufwand von einem Angreifer z.B. physikalisch manipuliert werden. Eine weitere Annahme sind Verletzungen der Sicherheitsziele durch Fehlbenutzung durch die Nutzer. Solch eine Fehlbenutzung liegt z.B. vor, wenn ein Benutzer sein Passwort leicht zugänglich irgendwo aufschreibt (z.B. gelbe Zettel am Rechner). Abschließend wird für alle bisher gesammelten Bedrohungen eine Schadensabschätzung gemacht. Dies ist interdisziplinär durchzuführen, da man über alle Disziplinen hinweg abschätzen muss, wie viel Schaden ein konkreter Angriff hervorrufen würde.
4. Sicherheitstests/Angriffssimulationen: Im Gegensatz zu der klassischen Sicherheitsanalyse, bei der Angriffe im Rahmen dieses Schrittes von Spezialisten simuliert werden, wird bei der interdisziplinären Sicherheitsanalyse eine Simulationsstudie mit Endbenutzern durchgeführt. Innerhalb dieser Studie verwenden Endbenutzer die zu analysierende Software bzw. das zu analysierende System unter realen Bedingungen. Dadurch können Angriffe identifiziert und durchgeführt werden, die nur unter Mitwirkung der Benutzer möglich sind. Andere Disziplinen können hierbei die Auswirkungen solcher Angriffe auf ihre Bereiche analysieren. Derlei Angriffe werden nicht zwingend wirklich durchgeführt, sondern es reicht diese so zu simulieren, dass sie für die Endbenutzer real erscheinen. Alle während der Simulation anfallenden Daten werden zentral gesammelt. Abschließend werden diese in Zusammenarbeit mit den anderen beteiligten Disziplinen ausgewertet. Auf Grundlage der gesammelten Daten wird abschließend bestimmt, wie wahrscheinlich die simulierten Angriffe tatsächlich in

der Realität durchgeführt werden könnten. Ein Beispiel wäre der simulierte Angriff auf ein Benutzerkonto (Kontoübernahme durch einen Angreifer). Die Wahrscheinlichkeit, dass dies möglich ist, hängt unter anderem von den kryptographischen Stärken der gewählten Endbenutzerpasswörter ab. Durch die gesammelten Daten kann festgestellt werden, wie viele Benutzer schwache Passwörter gewählt haben und dadurch kann die Wahrscheinlichkeit berechnet werden, mit der es möglich ist, ein Konto anzugreifen.

5. Schutzmaßnahmen: Auf Grundlage der im vierten Schritt bestimmten Wahrscheinlichkeiten wird eine Risikoanalyse durchgeführt. Jene Angriffe, die über einem bestimmten Wahrscheinlichkeitswert liegen, müssen erneut in Zusammenarbeit mit den anderen Disziplinen genauer betrachtet werden. Durch die Risikoanalyse können nun die Bedrohungen identifiziert werden, die minimiert werden müssen (d.h. hoher Schaden und hohe Wahrscheinlichkeit, dass der Angriff durchgeführt werden kann), die die bereits minimal sind (geringe Wahrscheinlichkeit, dass der Angriff durchgeführt werden kann, hohes Schadenspotenzial), und jene die irrelevant sind (geringes Schadenspotential und geringe Auftrittswahrscheinlichkeit).

Evaluation

In diesem Kapitel wird die neu entwickelte interdisziplinäre Sicherheitsanalyse angewendet, um den Evaluationsgegenstand MyGroup zu analysieren. Durch die Durchführung der Analyse werden einerseits Sicherheitserkenntnisse und Optimierungsmöglichkeiten für MyGroup gewonnen. Andererseits werden durch die Anwendung der interdisziplinären Sicherheitsanalyse ihre Vorteile gegenüber der klassischen Sicherheitsanalyse aufgezeigt. Innerhalb dieses Abschnitts wird dies an beispielhaft gezeigt.

Evaluationsgegenstand: MyGroup

Bei MyGroup handelt es sich um eine Web2.0-Anwendung, deren Ziel es ist, Mitglieder einer Arbeitsgruppe während ihrer täglichen Arbeit zu unterstützen. Mithilfe von MyGroup können Benutzer u. a. ihre Arbeitskollegen innerhalb der Arbeitsstätte mittels RFID-Tags, welche jeder Benutzer mit sich führt, lokalisieren, Gesprächswünsche untereinander austauschen, eigene soziale Kontakte verwalten sowie eigene Aktivitäten aus sozialen Netzwerken, z.B. BibSonomy und Twitter, der Arbeitsgruppe mitteilen.

Für die Durchführung der Sicherheitsanalyse wurden die wesentlichen Komponenten von MyGroup betrachtet, welche in Abbildung 1 dargestellt sind. Bei MyGroup müssen alle im Folgenden aufgezählten Komponenten betrachtet werden: Web-Schnittstelle, RFID-Schnittstelle, Datenbank, Applikationslogik, RFID-Tags, RFID-Reader

Neben den Komponenten sind für die Sicherheitsanalyse insbesondere die Kommunikationsschnittstellen relevant. Zum einen kommuniziert MyGroup mit RFID-Tags über eine entsprechende Schnittstelle, welche RFID-Reader nutzt. Zum anderen existiert eine Web-Schnittstelle (Kniewel & Schmidt, 2011), über die Benutzer mittels Webbrowser und HTTP mit der MyGroup-Webanwendung arbeiten können.

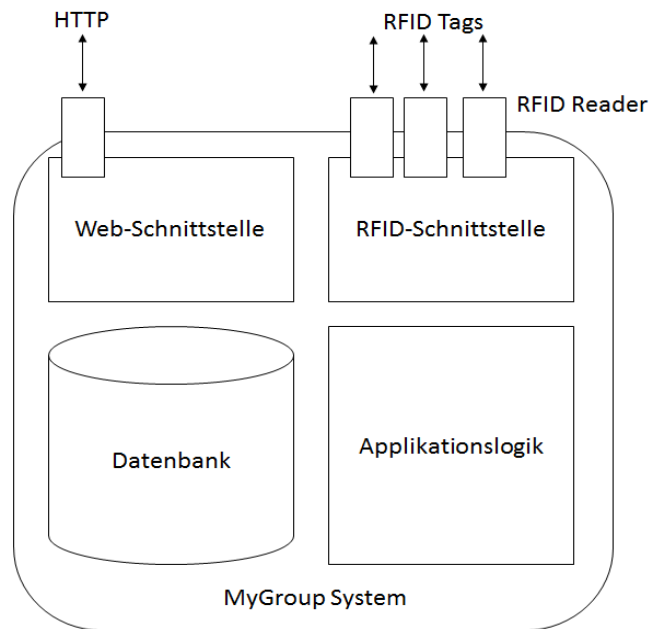


Abb.1: MyGroup-Komponenten

Beispielhafte Anwendung der Methode

Während der Anforderungsanalyse in Schritt 1 der interdisziplinären Sicherheitsanalyse wurden unter anderem die folgenden Sicherheitsziele festgelegt:

- Zugangskontrolle: Nur autorisierte Entitäten können auf Dienstleistungen und Informationen zugreifen.
- Vertraulichkeit: Informationen sind nur den befugten Entitäten zugänglich.

Bei der Bestandsaufnahme in Schritt 2 wurde bezüglich der Zugangskontrolle unter anderem folgendes festgestellt: Damit ein Benutzer MyGroup nutzen kann, muss dieser seinen Benutzernamen und ein Zugangspasswort angeben. MyGroup fordert an dieser Stelle ein Passwort, welches mindestens eine Länge von sechs Zeichen hat. Der Zugangsdialog der Webanwendung von MyGroup verhindert implizit durch den Einsatz eines Webframeworks einen Angriff mittels SQL-Injection. Zudem werden Passwörter innerhalb der MyGroup-Datenbank mittels MD-5 Hashverfahren abgelegt, allerdings wird hier kein Salt genutzt.

Bezüglich der Vertraulichkeit verhindert die Webanwendung von MyGroup, dass Benutzer die Daten anderer Nutzer einsehen können. Dies wird durch eine innerhalb der Webanwendung implementierten Rechteverwaltung erreicht. Die

Kommunikation zwischen Webbrowser der Benutzer und der MyGroup-Webanwendung basiert auf dem HTTP-Protokoll. Das Protokoll übermittelt Daten unverschlüsselt. Dadurch können alle Daten auf dem Netzwerkpfad zwischen Browser und dem Web-Server von Angreifern mitgelesen werden. Insbesondere werden die Zugangsdaten der Benutzer hierbei im Klartext übermittelt (Eckert, 2012).

Auf Grundlage der Bestandsaufnahme in Schritt 3 wurden im Sinne der klassischen Sicherheitsanalyse folgende Schwachstellen identifiziert. Eine Passwortlänge von mindestens sechs Zeichen entspricht einer möglichen Mindeststärke von 39 Bit. Dies ist heutzutage zu gering, da ein einzelner Computer 39 Bit in wenigen Stunden vollständig durchsuchen kann (Ferguson, 2012). Über die Webschnittstelle ist das Durchsuchen deutlich langsamer, da das Netzwerk die Geschwindigkeit begrenzt. Wenn ein Angreifer allerdings Zugriff auf die gehashten Passwörter hat, so ist dies eine ernste Bedrohung. Diese Schwachstelle wird noch dadurch verstärkt, dass die Benutzer nicht gezwungen sind, Sonderzeichen und Zahlen zu verwenden, wodurch die tatsächliche Bitstärke des Passworts deutlich geringer sein kann und anstelle einer vollständigen Suche ein Wörterbuchangriff möglich ist. Das Risiko bzw. die Schadensabschätzung eines korrumpierten Benutzerkontos ist somit gegeben und als sehr hoch einzustufen, da Benutzer, deren Benutzerkonto übernommen wurde, mit hoher Wahrscheinlichkeit die Nutzung von MyGroup beenden. Eine weitere Schwachstelle stellt die Nutzung des HTTP-Protokolls ohne weitere Verschlüsselung dar. Da die vollständige Kommunikation von Angreifern mitgelesen werden kann, ist es diesen möglich, alle Benutzerdaten, auch Passwörter, zu erlangen. Auch hier ist das Risiko als hoch einzustufen, da persönliche Daten, insbesondere Passwörter der Benutzer, an Dritte gelangen und somit auch eine Korrumpierung des Benutzerkontos ermöglichen.

Für die interdisziplinäre Sicherheitsanalyse wurden weitere Schwachstellen angenommen, für die keine technischen Voraussetzungen gegeben sind. Angenommen ein Angreifer erlangt Zugriff auf die vollständige Datenbank und erhält so Zugriff auf alle Benutzerdaten inklusive der gehashten Passwörter. Aufgrund der oben aufgeführten Schwächen wäre es ihm möglich, einen Teil dieser zu brechen. Innerhalb der Simulationsstudie wurde überprüft, wie viele Simulationsteilnehmer schwache Passwörter verwendeten und dadurch gebrochen werden konnten. Hierbei konnten 73,91 % aller Passwörter aus deren Hash-Werten errechnet werden. Die Angriffswahrscheinlichkeit ist entsprechend der Anzahl der ermittelten Passwörter somit hoch einzuschätzen.

Ein weiterer, innerhalb der interdisziplinären Sicherheitsanalyse durchgeführter, Angriff bestand aus einem auf Social-Engineering basierenden Ansatz. So wurden alle Teilnehmer der Simulationsstudie von einem dritten Unbekannten, welcher sich als Administrator ausgab, aufgefordert, ihr Passwort auf ein einfaches vorgegebenes Passwort („123456“) abzuändern. Hierbei konnten 66,67 % der erreichten Teilnehmer tatsächlich überzeugt werden, das Passwort zu ändern. Durch die Änderung des Passwortes auf das Vorgegebene hätte ein Angreifer das Konto des Be-

nutzers jetzt nutzen können. Die Angriffswahrscheinlichkeit ist entsprechend der Anzahl, der von den Benutzern freiwillig geänderten Passwörter, somit als hoch einzuschätzen.

Außerdem wurden die Teilnehmer der Simulationsstudie aufgefordert, die MyGroup-Webanwendung über eine manipulierte URL aufzurufen. Diese Aufforderung wurde durch eine gefälschte Email an alle Teilnehmer gesendet (Phishing). Hierbei konnten 100 % der Teilnehmer davon überzeugt werden, die manipulierte URL zu nutzen. Durch die Nutzung der manipulierte URL hätte ein Angreifer jetzt den vollständigen Netzwerkverkehr einsehen und lesen können. Die Angriffswahrscheinlichkeit ist somit als hoch einzuschätzen.

In Schritt 5 der interdisziplinären Sicherheitsanalyse wurden in Zusammenarbeit mit den anderen beteiligten Disziplinen Schutzmaßnahmen gegen die identifizierten Schwachstellen entwickelt. Um das Risiko von korrumpierten Benutzerkonten trotz aller identifizierten Schwachstellen zu minimieren, sollten Passwörter in der Datenbank mit Salt-Werten und einer besseren Hash-Funktion gespeichert werden. Außerdem sollten Benutzer von der MyGroup-Webanwendung angehalten werden, stärkere Passwörter zu verwenden. Die konkrete Umsetzung von stärkeren Passwörtern sollte interdisziplinär mit der Usability-Disziplin erarbeitet werden.

Um die Vertraulichkeit zu gewährleisten, sollte das HTTP-Protokoll durch das sichere HTTPS-Protokoll ersetzt werden, wodurch der gesamte Netzwerkverkehr verschlüsselt wäre und ein Mitlauschen unmöglich ist (Schwenk, 2010). Zusammen mit der Disziplin Recht könnte ein rechtlicher Rahmen erarbeitet werden, der vorschreibt, dass immer dann, wenn persönliche Daten über ein öffentliches Medium übermittelt werden, diese zu verschlüsseln sind.

Um Phishing erfolgreich entgegen zu wirken, könnte zusammen mit der Disziplin Recht ein rechtlicher Rahmen geschaffen werden, der bestimmte Aufforderungen zur Änderung von Benutzerdaten gesetzlich verbietet, wie z.B. die Änderung eines Passworts auf einen bestimmten Wert. Dadurch könnten Benutzer sicher sein, dass solche Aufforderungen niemals von legitimen Quellen kommen und diese zu ignorieren sind. Zusätzlich sollte zusammen mit den Disziplinen Usability und Recht, darüber aufgeklärt werden, dass bestimmte Informationen niemals von legitimen Quellen abgefragt werden würden (wie z.B. bei Banken).

Zusammenfassung

In diesem Beitrag haben wir einen neuen Ansatz für eine interdisziplinäre Sicherheitsanalyse für eine ubiquitäre Anwendung vorgestellt. Am Beispiel der im VENUS Projekt entwickelten Anwendung MyGroup wurde aufgezeigt, wie eine solche interdisziplinäre Sicherheitsanalyse in Zusammenarbeit mit weiteren Disziplinen, wie z.B. Recht oder Usability, innerhalb einer Simulationsstudie durchgeführt werden kann. Innerhalb der Simulationsstudie wurden sowohl technische als auch nicht-technische Bedrohungen mittels der Analyse identifiziert. Die Ergebnisse

dieser Sicherheitsanalyse wurden genutzt, um in Zusammenarbeit mit den anderen Disziplinen geeignete Schutzmaßnahmen zu entwickeln. Diese waren z.B. aus Sicht der IT-Sicherheit die Anpassung von Sicherheitstechniken, aus Sicht des Rechts die Anpassung des rechtlichen Rahmens und aus Sicht der Usability die Anpassung der Benutzerschnittstellen.

Literatur

- Eckert, C. (2012). *IT-Sicherheit: Konzepte, Verfahren, Protokolle*. München: Oldenbourg Verlag
- Ferguson, N., Schneier, B., & Kohno, T. (2012). *Cryptography Engineering: Design Principles and Practical Applications*. Hoboken: John Wiley & Sons
- Kniewel, R., & Schmidt, L. (2011). Das Design ubiquitärer Systeme am Beispiel von MyGroup. In: Schmid, S.; Elepfandt, M.; Adenauer, J. & Lichtenstein, A. (Hrsg.): *Reflexionen und Visionen der Mensch-Maschine-Interaktion - Aus der Vergangenheit lernen, Zukunft gestalten: 9. Berliner Werkstatt Mensch-Maschine-Systeme* (Berlin 2011). Düsseldorf: VDI-Verlag, 2011, 609-614.
- Paulus, S. (2011). *Basiswissen Sichere Software*. Heidelberg: dpunkt.verlag GmbH
- Schwenk, J. (2010). *Sicherheit und Kryptographie im Internet: Von sicherer E-Mail bis zu IP-Verschlüsselung*. Vieweg+Teubner Verlag.

The SHERPA driving simulator as a technical platform for driver's state monitoring studies

Philippe Simon¹, Jérôme Floris¹ and Jean-Christophe Popieul¹

Keywords: Car driving, simulation, driver's state, monitoring and warning device

Abstract

Several studies were conducted at the laboratory some years ago about drowsiness detection and mental workload assessment in the context of car driving. These studies were made in a driving simulator and involved a posteriori data analyses, with the aim to identify symptoms of potentially dangerous situations/states. More recently, in the continuity of these works, developments have been made in the SHERPA simulator that equips the laboratory to implement similar analyses but in real time, while driving. A demonstrator has been developed, called D2M for "Driver and Driving Monitoring", in the form of a warning system combining a Lane Departure Warning, a Headway Collision Warning, a distraction warning and drowsiness warning. Rather than a prototype, this system is intended as a technical platform allowing studies on several issues: sensors and data collection, data analysis and HMI to inform the driver about his own state.

Introduction

For several decades, many works are dedicated to the estimation of the "driver's state". They are generally designed to detect potentially hazardous states both for the driver and for the road safety: high mental workload (Schaap et al, 2009), visual and cognitive distraction (Liang et al, 2010; Young et al, 2011), inattention (Regan et al, 2011), drowsiness or sleepiness (Daza et al, 2011; May et al, 2009), etc. Most often, the articles present results obtained in experiments conducted in a simulator or in a real vehicle, data being analyzed a posteriori in order to identify reliable and sensitive symptoms of these hazardous states. Fewer works are dedicated to a real-time analysis, while driving.

"Real-time" techniques are of course essential to develop warning systems such as those proposed by car manufacturers or automotive suppliers. Moreover, new needs have arisen with recent works on ADAS (Blaschke et al, 2009) and more especially on systems providing a "shared control" between the driver and an automated system (Carsten et al, 2010; Sentouh et al, 2010). The assessment of the driver's state is then essential and can be used in several ways: as a criterion for the task sharing between the driver and the automated controller (e.g., in order to "smooth" the driver's workload), as a way of checking the effective involvement of

¹ LAMIH, CNRS UMR 8201, Université de Valenciennes et du Hainaut Cambrésis, France

the driver into the control loop (e.g., in order to keep a good "situation awareness"), or as an ultimate protection that can launch an automated emergency procedure (e.g., emergency stop in the case the driver would fall asleep). Dealing with automated driving, it should be kept in mind that the "Vienna convention" (dating from 1968 but still valid) states that "the driver must constantly be in control of his vehicle" (sic).

To support its works about automated driving, the laboratory has launched further works about real-time assessment of the driver's state (Girard et al., 2010). A demonstrator called D2M for "Driver & Driving Monitoring" was developed in the SHERPA driving simulator that equips the laboratory. The D2M has the form of a warning device in case of distraction or drowsiness. It combines a "Lane Departure Warning" (LDW), a "Forward Collision Warning" (FCW) and a driver monitoring system based on eye blinks and head/eye movements. Beyond its role as a demonstrator, the D2M is intended as a technical platform dedicated to experimental studies and developments about driver's state assessment. The goal is to provide researchers and engineers with a tool that facilitate works on central issues such as:

- Which data/indicators in order to characterize the driver's state?
- What real-time processings?
- How to inform the driver about the diagnosis made on his own "state"?
- How to ensure the acceptability of such systems?

The present paper provides a technical description of the D2M.

SHERPA driving simulator

SHERPA-lamih (Figure 1) is a copy of the driving simulator developed by the French car manufacturer PSA Peugeot Citroën. This simulator is based on a Peugeot 206 mock-up fixed on a 6 degree-of-freedom motion system (Rexroth Hydraudyne). The front view is displayed on flat screens via five video projectors (ProjectionDesign DLP), allowing an overall field of view of 240°. The three rear-view mirrors are equipped with small LCD monitors that provide a realistic rear field of view. A sound generator reproduces the sound of the engine, aerodynamic noise, as well as the sounds coming from the other vehicles on the road. From a hardware point of view, SHERPA is built around a fiber-optic ring that connects together 15 PC computers, each one dedicated to a specific task: dynamic model calculation, image generation, sound generation, etc. Such a configuration provides a good realism level, compatible with the studies conducted at the lab.

The distributed hardware architecture of SHERPA (Figure 2) facilitates the prototyping of ADAS / IVIS. Indeed, each "node" connected to the fiber-optic ring of SHERPA is equipped with a "reflective memory board". The same memory mapping is used by all nodes. The operating principle is simple: the modification of any memory data on one of the nodes is "instantly" passed on to the other nodes connected to the fiber-optic. Data exchanges are therefore easily done by memory

sharing. This kind of communication allows, despite a massively distributed architecture, to satisfy the real-time constraints inherent to any interactive simulation application. To prototype a system such as the D2M, a new node has been added to the ring (a PC computer equipped with a reflective memory board) which allows to easily access to all the simulation variables: driver's actions, kinematics and dynamics of the vehicle, data linked to the traffic's vehicles, etc.



Fig. 1: SHERPA-lamir driving simulator.

For the additional instrumentation (e.g., physiological sensors), the chosen solution is to use a specific node in charge of collecting data flows (via a TCP / IP connection, a RS232 link, A/DIO boards, etc.) and of copying them into the reflexive memory board, making them available to all nodes connected to the ring. Regarding the D2M, the application uses the data provided by the faceLAB system installed into the SHERPA simulator.

More and more sensors feature a CAN interface that is commonly used in the automotive domain. So a CAN bus has been installed in the simulator's mock-up. E.g., a torque sensor is connected to this CAN bus to accurately know the torque applied to the steering by the driver. Here again, the access to the data produced by CAN sensors is facilitated as they are directly available in the reflective memory boards. A specific node is used as a "gateway" between the CAN bus on one hand, and the simulator's ring on the other hand. This node runs with the CANoe development environment (Vector company) that offers multiple possibilities in terms of prototyping, testing and also simulation of the components.

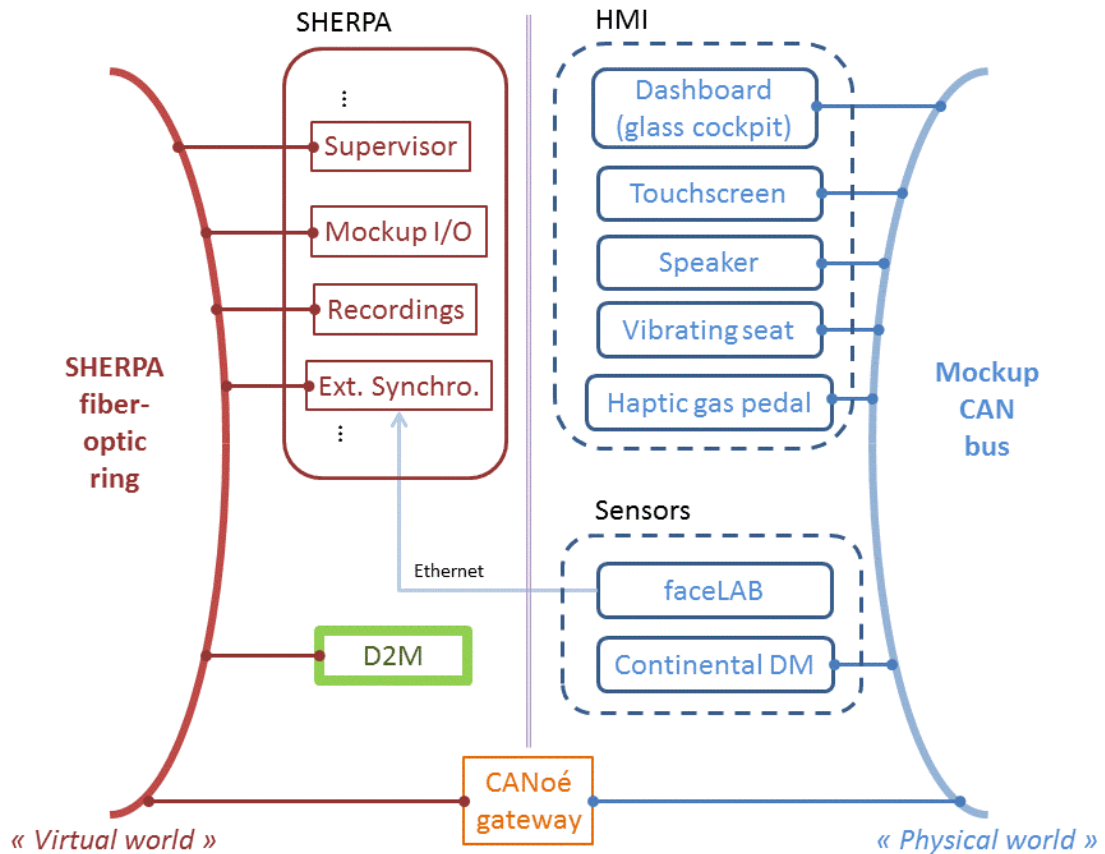


Fig. 2: Hardware architecture of SHERPA.

Operating principles of the D2M

By definition, the driver's state is a notion intrinsic to the driver. This is why physiological indicators (especially EEG) are often used as reference indicators (Brookhuis et al, 2010). In the literature, one often speaks of "direct methods" to describe those using data measured on the driver himself. The problem is that these methods are often "invasive" for the driver (except perhaps those based on video recording e.g., for the analysis of eye blinks) and they are therefore unrealistic in real life driving.

Other methods, described as "indirect methods", aim to analyze the actions of the driver on the vehicle or the vehicle's behavior on the road. These methods are transparent to the driver and are obviously favored by car manufacturers and automotive suppliers.

Many studies underline the importance of mixing the two types of methods and "merge" the two types of indicators (Sandberg et al, 2009; Daza, 2011). Intuitively, one can guess the value of such a merging, which allows to detect both a physiological changes on the driver and also a consequence on his behavior or on the vehicle's behavior. This approach should allow a more robust diagnosis, hence its use for D2M system.

As seen previously, all data (car data and driver data) are available in the reflective memory, synchronization problems being so solved. In this first version of the D2M, processings are implemented in a Matlab / Simulink module.

Data processing is carried out in four main parallel branches:

- The "LDW" branch that deals with the lateral position of the vehicle on the road,
- the "FCW" branch which calculates the time to collision,
- the "drowsiness" branch based on the PERCLOS indicator provided by the faceLAB system,
- the "Distraction" branch that focuses, upon a 1 minute sliding window, on the "off road" gaze diverting time; the D2M system can work with eye movements or head movements, the latter being easier to measure (no calibration needed) and having a lower dynamic thus inducing fewer "false alarms".

Each of the four branches generates independant alerts to the driver via a dedicated screen and also via a speaker. These individual alerts constitute a first alert level, visually colored in orange and associated with short sounds. In the simultaneous presence of two kinds of warning – e.g., LDW alert associated with PERCLOS alert - a second alert level is raised, visually colored in red and associated with longer sounds lasting all the time of the alarm. Not to mention data fusion, this is a first attempt to take into account simultaneously both car data and driver data for the diagnostic.

HMI and first tests

To facilitate prototyping of ADAS / IVIS, SHERPA is equipped with a "glass cockpit" composed of two LCD screens for the dashboard and a touch screen in central position for HMI purposes. The content of the screens is described using XML files for more flexibility (no programming needed). Data exchanges with the simulation is done via the CAN bus and the CANoe gateway.

Figure 3a shows the faceLAB system permanently installed in the SHERPA's mock-up. It also shows the touch screen used for prototyping HMI. Figure 3b shows the HMI developed specifically for the D2M. Simple graphical representations are used to inform the driver of unintentional white line crossing, of an impending collision, of a drowsiness warning or of a distraction warning. This central screen (actually an EEEPC) also manages audio messages inside the cockpit.

The HMI also uses CAN components such as a "haptic" gas pedal (developed by the Continental company) or a vibratory system installed into the driver's seat (laboratory development). These systems are controlled simply through orders sent to the reflective memory board. The same board also allows to easily implement more "original" solutions such as e.g., applying a slight oscillation on brakes to alert the driver or making a "suggestion of action" on the steering wheel.



Fig. 3: D2M prototype installed in the SHERPA simulator: faceLAB system (3a) and dedicated HMI (3b).

Figure 2 points out a significant advantage of the D2M i.e., sensors aspects, data processing and HMI are completely independent, which naturally facilitates development and testing.

This first version of the D2M has been evaluated with a small population of 6 drivers. The evaluation focused on distraction and drowsiness aspects. The course was done in five steps: 1) Brief presentation of the system, 2) driving with an added telephoning task in order to cause distractions, 3) driving with simulated episodes of drowsiness (actually triggered by the experimenter after a "briefing" of the driver), 4) filling out a questionnaire, 5) open discussion with the experimenter. The questionnaire consisted of 16 questions dealing with the usability of the system, the understanding of its operation, the driver acceptance, and finally its performance and perceived usefulness. This evaluation is of course a very first step. But the information gathered was used to correct some drawbacks and led to a second version under development.

Conclusion and prospects

The paper presents the D2M system developed in the SHERPA driving simulator. This system is not a prototype, but a technical platform dedicated to studies about the estimation of the driver's state. The hardware and software architecture facilitates using of numerous data (both car data and driver data) and also rapid prototyping of new algorithms for data processing. The development of this type of system is usually done by successive iterations (Mulder et al, 2010). Hence the importance of using software tools oriented to rapid prototyping, such as Matlab / Simulink for data processing or CANoe for the implementation of CAN components.

A HMI has been developed and tested. These first tests concluded that the algorithms for data processing need to be improved, as well as the way of generating alerts for the driver. Other possible improvements were identified, such as:

- Better alerts modulation based on the precise driving context,
- customization of alert levels according to the preferences of the driver and / or his driving style,
- customization of modalities used for alerts (visual / audio / haptic).

Dealing with the technical aspects, the next version will be based on the RTMaps soft-ware environment (Intempora company) with a double advantage: a large number of recognized sensors (physiological sensors, eye trackers, ...) and also the existence of image processing libraries dedicated to driving. The data processing algorithms will be still developed with Matlab / Simulink.

Further experiments will be of course conducted to assess the performance of the system. In addition, we will look closely at acceptability and trust aspects, which are key issues for a successful deployment of this kind of alert device.

Acknowledgments

The present research work has been supported by International Campus on Safety and Intermodality in Transportation, the Nord Pas-de-Calais Region, the European Community, the Regional Delegation for Research and Technology, the Ministry of Higher Education and Research, and the National Center for Scientific Research. The authors gratefully acknowledge the support of these institutions. The authors also wish to thank PSA Peugeot Citroën for its technical support with the SHERPA driving simulator.

References

- Blaschke, C., Färbe, B., Limbacher, R., Trefflich, B., Breyer, F., Freyer, J., & Mayer, S. (2009, Sept. 28-29). *Online estimation of the driver's state. Enhancement of lane-keeping assistance*. Paper presented at the first international conference on Driver Distraction and Inattention, Göteborg, Sweden.
- Brookhuis, K.A., & De Waard, D. (2010). Monitoring drivers' mental workload in driving simulator using physiological measures. *Accident Analysis and Prevention*, Vol. 42, 898-903.
- Carsten, O., Lai, F., & Barnard, Y. (2010, Aug. 31–Sept. 3). *Semi-automated driving: How does the supported task affect driver response?* Paper presented at the 11th IFAC international conference on Analysis, Design, and Evaluation of Human-Machine Systems, Valenciennes, France.
- Daza, I.G., Hernandez, N., Bergasa, L.M., Parra, I., Yebes, J.J., Gavilan, M., Quintero, R., Florca, D.F., & Sotelo, M.A. (2011, Oct. 5-7). *Drowsiness monitoring based on driver and driving data fusion*. Paper presented at the 14th international IEEE conference on Intelligent Transportation Systems, Washington, DC, USA.

- Girard, J-M., Younsi, K., Floris, J., Simon, P., Loslever, P., & Popieul, J-C. (2010). An analysis of driving performance evolution under high workload conditions. *Advances in Transportation Studies an International Journal*, Section B 20, 49-58.
- Liang, Y., & Lee, J.D. (2010). Combining cognitive and visual distraction: less than the sum of its parts. *Accident Analysis and Prevention*, Vol. 42, 881-890.
- May, J.F., & Baldwin, C.L. (2009). Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transportation Research Part F*, doi:10.1016/j.trf.2008.11.005.
- Mulder, M., Abbink, A., Van Passen, M.M., & Mulder, M. (2010, Aug. 31–Sept. 3). *Haptic gas pedal support during visually distracted car following*. Paper presented at the 11th IFAC international conference on Analysis, Design, and Evaluation of Human-Machine Systems, Valenciennes, France.
- Regan, M.A., Hallett, C., & Gordon, C.P. (2011). Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accident Analysis and Prevention*, Vol. 43, 1771-1781.
- Sandberg, D., Wahde, M., Anund, A., Kecklund, G., & Akerstedt, T. (2009, Sept. 28-29). *Detecting sleepiness in truck drivers*. Paper presented at the first international conference on Driver Distraction and Inattention, Göteborg, Sweden.
- Schaap, T.W., Van der Horst, A.R.A., Van Arem, B., & Brookhuis, K.A. (2009, Sept. 28-29). *The relationship between driver distraction and mental workload*. Paper presented at the first international conference on Driver Distraction and Inattention, Göteborg, Sweden.
- Sentouh, C., Debernard, S., Popieul, J-C., & Vanderhaegen, F. (2010, Aug. 31–Sept. 3). *Toward a shared lateral control between driver and steering assist controller*. Paper presented at the 11th IFAC international conference on Analysis, Design, and Evaluation of Human-Machine Systems, Valenciennes, France.
- Young, K.L., & Salmon, P.M. (2011). Examining the relationship between driver distraction and driving errors: A discussion of theory, studies and methods. *Safety Science*, doi:10.1016/j.ssci.2011.07.008.

Probabilistic Techniques to Diagnose Human Operator State

Nedjemi Rachedi¹, Denis Berdjag¹, Frédéric Vanderhaegen¹

Keywords: Human-Machine System; Hidden Markov Models; Bayesian Network; State diagnosis.

Abstract

Human's states and behaviors are one of factors that must be monitored to ensure the safety of Human-Machine systems. This article is concerned with driver state diagnosis in rail systems from endogenous and exogenous Human-Machine data such as driving behaviors. The references of these behaviors are modeled with Hidden Markov Models. The observed behaviors are recognized by comparison with reference behaviors. Other parameters and measures are used with the precedent result through data fusion techniques by a Bayesian Network in order to detect three states of driver: normal, drowsy and over-reactive. A reliable and nonintrusive detection system is presented and experimental study is proposed in illustration. Future research directions are discussed.

Introduction

Human-machine systems safety enhancement is one of the important goals in transportation system. The studies show that the risk of human errors increases with the vigilance decline; it is even one of the main causes of transport accidents. Devices such as « dead-man » used in rail transport systems nowadays to reduce the risk, do not detect the states of the driver but especially his loss of consciousness. Besides, these devices are few accepted by drivers [6].

Many challenges are present in driver's states modeling and monitoring. Initially, these states are not directly observables but could only be inferred from observable data. Moreover, the observations are generally uncertain, incomplete and heterogeneous. Driver's state (drowsiness, fatigue and inattention) detection approaches are classified according to the type of used data:

- The direct data [3, 17], including driver biological measures (e.g., EEG, ECG) and driver physical measures (e.g., blinking activity, head movement).
- The indirect data [8, 9], also named driving performance measures (e.g., acceleration, steering wheel angle).

¹ LAMIH

The University of Valenciennes and Hainaut-Cambresis
Valenciennes, France

nedjemi.rachedi;denis.berdjag;frederic.vanderhaegen@univ-valenciennes.fr

- The subjective report measures (e.g., Epworth or Karolinska sleepiness scale).

These data allow refining the analysis of performance and vigilance [14, 25].

There is also a category of approaches which uses hybrid measure, i.e., combine the direct and indirect measures [12].

The sensors measuring the biological activity are reliable [5], but are intrusive and relatively cumbersome. To conceive a reliable and efficient system of detection, it is recommended to use both the direct and indirect measures. That allows reducing false alarms and improving the rate detection [5].

Our work aims to develop a reliable and nonintrusive detection system, able to interpret the driving behaviors and to deduce the driver's state. For that purpose, a driver's behaviors detection approach is proposed. Based on Hidden Markov Models construction (HMM), the method uses indirect measures to distinguish normal driving behaviors from the risky ones, focusing on drowsiness and aggressive behaviors. The behavior detection result is then fused with two direct measures and one subjective measure, by using a Bayesian network. The global detection system infers the driver's state according to the "Fig.1".

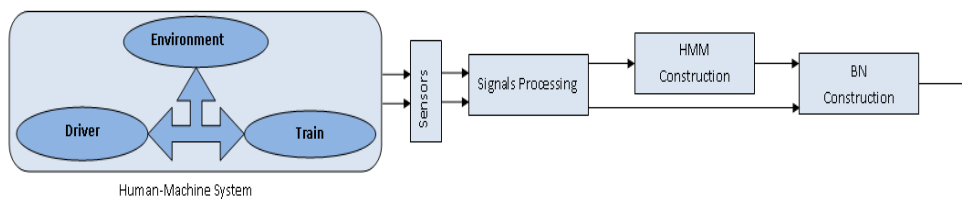


Figure 1. State detection Methodology

The paper is organized as follows. Section 2 presents the diagnosis system. The theoretical aspects of the HMM and BN and their applications in our context are described in the section 3. The next section shows the experimental results. Finally, in the last section, a conclusion and some perspectives are presented.

System diagnosis presentation

A several methods have been presented in the literature, for modeling data and fusing information to monitor driver's state. These methods include probabilistic methods, evidential reasoning, statistical learning, fuzzy theories, and neural networks. Among the various methods, we use probabilistic methods, because they allow heterogeneous data fusion and uncertainties modeling at different levels of abstraction. The studied Human-Machine system is described by the triplet Train-Driver-Environment.

The notable change/improvement that we propose, compared to the existing works, is in considering direct, indirect measures and also subjective information about the human operator character. Besides, our approach is not limited to the

drowsiness detection, but also detects another risky state, named over-reactive state i.e., aggressive driving.

The kinematics of the train depend on the driver's decisions and attitudes (e.g., acceleration, braking). These decisions are conditioned by intrinsic and/or extrinsic events such as the environment changes (e.g., intersection, halt), and also by the driver's emotional and physiological states (e.g., drowsiness, fatigue.). So, we propose to detect a behavior to deduce afterward a state. To detect the behavior, the method consists in observing the train speed evolution, and comparing it to typical evolutions characterizing known driving behaviors.

Our approach is inspired by the article [7], where the authors use the vehicle dynamics to detect the driver's actions in an intersection. Otherwise, we are not limited by the behaviors detection but we also deduce the driver's state. The proposed methodology deals with a rail transport system however, it can be easily used in other human-machine system in transport domain.

The driver is modeled by a 2nd order transfer function in cascade with a proportional integral controller and a delay, given by:

$$\left(\frac{K}{S^2 + 2\zeta\omega_n S + \omega_n^2} \right) (K_p + K_i / S) e^{-\tau S} \quad (1)$$

An equivalent modeling of human operator in aeronautics was proposed by [15, 23]. The driver sends an acceleration which will be translated into terms of speed by the train. Various speed profiles are obtained by identification of the transfer function's parameters. Each driving behavior is associated to a specific set of speed profiles. We retain three profile sets, i.e., three driving behaviors. Each set is presented by a typical profile in "Fig. 2".

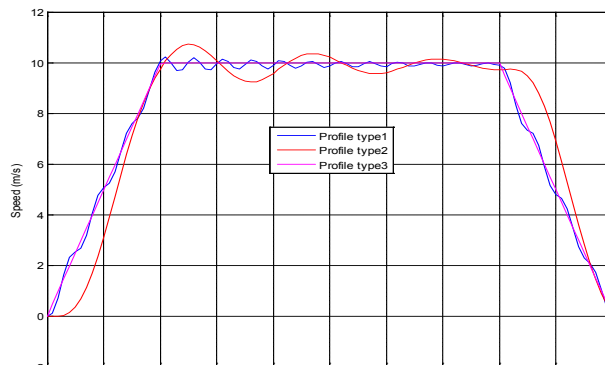


Figure 2. Speed evolutions for each driver's behavior

The profiles are obtained from a study scenario consisting in driving the train on a bounded straight section, where the driver has to follow at best a trapezoidal speed profile. All the variables are normalized “Fig. 2”.

The driver, characterized by the Profile type1, tends to overreact by trying to respect the prescription. He tries to correct his speed exceeding in a rather fast way by showing certain impatience. The driver, characterized by the profile type 2, tends to deviate from the prescribed speed, he will try to correct this exceeding, but that will often be with a more or less important delay. The driver, characterized by the profile type 3, respects the speed reference perfectly: that will be the normal driving profile.

The Hidden Markov Models (HMM) are used to model these driving behaviors. Each HMM parameters are determined by learning from a set containing train speed measures acquired during the study scenario. This learning is made offline, contrary to the driving behavior detection done online. The speed measures are sampled; afterward a features extraction is done. The extraction method is based on the calculation of the angle's sine between the horizontal axis and the straight line passing by the previous and current point. The obtained angle's sine is rounded with a precision of 0.1. This method is inspired by the handwriting recognition methods [21]. The features extracted from the speed profiles constitute the emission parameters of the HMM. This emission set contains integers between 1 and 20.

The same processing is performed on the measured data during driving, in order to recognize any driver behavior. The obtained sequence is then compared with the emissions of HMM (associated with the three basic behaviors). In this comparison, the belonging probability of the observed speed profile to the three typical profiles is calculated. However, the error possibility remains in certain cases where the probabilities are close. For the detection reliability, three additional data are used:

- The first is information about the personality of the driver, which is his degree of sensations seeking.
- The second is a direct data: the driver's reaction time to a secondary task called psychomotor vigilance task. The stimulus could be visual or auditory.
- The third is the ocular activity measure of human operator. This measure is useful for characterizing the driver's state and some parameters such as eye blinking are revealers of reduced alertness [22]. For this study the following parameters are considered: frequency and duration of eyes blinking. A Bayesian network is used for the data fusion.

Theoretical aspects and application

Hidden Markov Models

The use of the HMM is justified by their capacity to represent problems where states are not directly observable. Besides, an HMM is a stochastic model, what is suited to the nondeterministic nature of the drivers behavior.

An HMM is a probabilistic automata defined by the quintuplet $\Lambda = (\Sigma, \Omega, \Pi, A, B)$ where:

- The vector (Π) contains start probabilities.
- The set (Σ) contains hidden states.
- (A) is the matrix of transition probabilities between states.
- The set (Ω) contains the emission symbols.
- (B) is the matrix of emission probabilities.

An HMM $\Lambda_i / i=1,2,3$ is associated with each speed profile set. Each HMM contains five hidden states and 20 emissions symbols "Fig.3.a". To determine the HMM parameters, the algorithm « Baum-Welch » is used in the learning step [18]. The driving behavior detection consists in identifying the HMM, the most likely to generate the observation sequence obtained from the speed profile "Fig. 3.b".

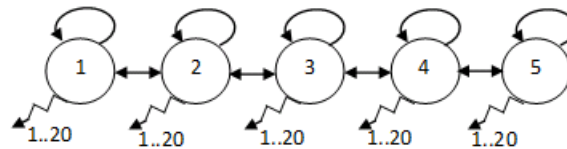


Figure 3.a

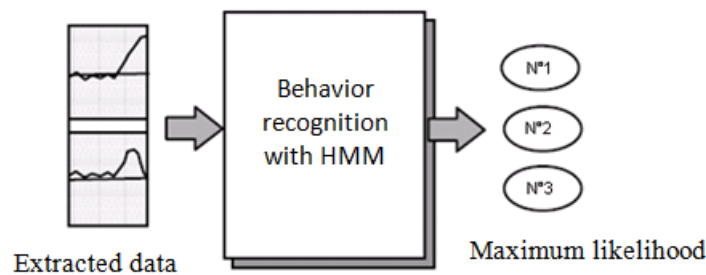


Figure 3.b

Figure 3. Hidden Markov Model structure and use

The probabilities estimation is done using "Forward-Backward" algorithm [18].

Bayesian Network

The reasons of using Bayesian Network (BN) are the heterogeneous nature of data and the uncertainties measures related to the dynamic environment in which the system Man-Machine evolves.

The Bayesian networks are models of knowledge representation combining the probability approaches and the graphs theory. The BN are used in the systems diagnosis, data analysis and fusion, the decision support and control systems [13]. BN were also used in human fatigue monitoring [16].

A Bayesian Network is a probabilistic graphical model defined by [8] as:

- a directed acyclic graph $G, G=(V, E)$, where V is the set of nodes of G , and E is the set of edges of G ;
- a finite probabilised space (W, Z, P) ;
- a set of random variables associated to the nodes of the graph and defined on (W, Z, P) , such that: $P(V_1, V_2, \dots, V_n) = \prod_{i=1}^n P(V_i | C(V_i))$ where $C(V_i)$ is the set of causes (parents) of V_i on the graph G .

The proposed BN "Fig.4" has four parent nodes, corresponding to the four data:

- Driver Behavior: it is the probability that the driver will have one of the three typical behaviors (calculated previously from the HMM).
- Driver reaction: This is the driver reaction time, measured, normalized and classified according to three categories:
 - 1. Small if the value is in $[0 \ 0.2]$;
 - 2. Medium if the value is in $[0.2 \ 0.7]$;
 - 3. Large if the value is in $[0.7 \ 1]$.
- Driver Character: This is the estimated driver degree of "Sensation-seeking" classified following three categories: Small, Medium and Large.
- The driver ocular movements: this data is recorded, normalized and classified according to three categories:
 - Small: if the value is in $[0 \ 0.3]$, corresponding to a small blinking frequency of short duration.
 - Medium: if the value is in $[0.3 \ 0.7]$, corresponding to a small blinking frequency of average duration.
 - Large: if the value is in $[0.7 \ 1]$, corresponding to a large blinking frequency of long duration.

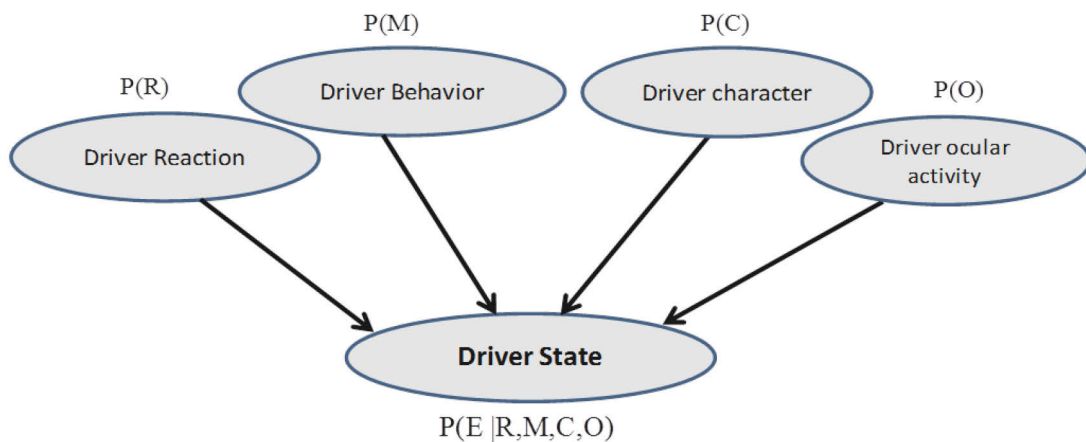


Figure 4. Bayesian Network

The node "Driver State" includes three states: "Over-reactive", "Drowsy" and "Normal." We consider two risky states "Over-reactive" and "Drowsy" because

their occurrence causes an increase in the accident probability. To reduce the diagnosis complexity, risky states detection may be considered only if the "normal" state is detected with a probability less than 50%.

During the construction of the conditional probability table, we considered that the driver ocular activity is preponderant in the driver state characterization. The respective influence order of the other nodes is defined as follows: "driver behavior", "driver reaction" and "driver character".

We give two situations in the form of rules as an example:

- IF driver behavior is "normal" AND ocular movement is "large" AND driver reaction is "large" AND driver character is "large" THEN driver state is "Drowsy".
- IF driver behavior is "over-reactive" AND ocular movement is "medium" AND driver reaction is "small" AND driver character is "small" THEN driver state is "Over-reactive".

After estimating the probability of the driver behavior, recording and processing the other data, all information are instantiated in the BN. So, the driver's state is evaluated and expressed in terms of probabilities.

Experimental results

The experiment was carried out on the railroad simulation platform COR&GEST "Fig.5". The platform includes a reduced model of railroad network "Fig.5.a", coupled with supervision and driving interfaces. The driving interface is shown in the "Fig. 5.b".

The driving behavior is classified by using the recoded speed profiles and estimated by the HMM. Using the proposed BN, a multi-sensory fusion is made to detect the driver's state. The methodology of detection was tested in simulation in [19].

The driving interface is equipped with a Tobii oculometer. It delivers in real time variables related to the driver's visual activity such as: pupil diameter, eyes movement in the horizontal/vertical axis, fixation duration.

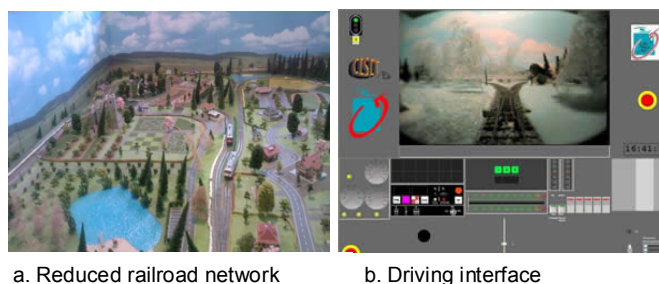


Figure 5. COR&GEST platform

The experiment was conducted on non-expert subjects. During the driving scenario, of 15 minutes, the speed limitation and the stop in station must be respected.

We show the evolution of the train speed during the driving scenario for subjects n°1 and n°2 with speed limits along the path “Fig.6”.

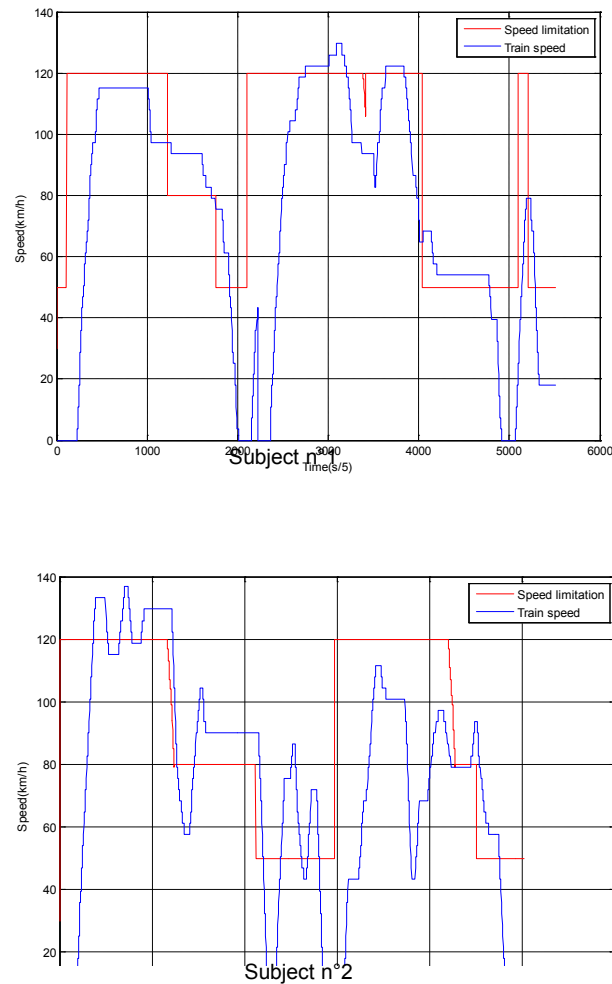


Figure 6. Train and limitation speed evolution

The results of behaviors detection using HMM is showed in Fig 7. The detection window is of 10 seconds.

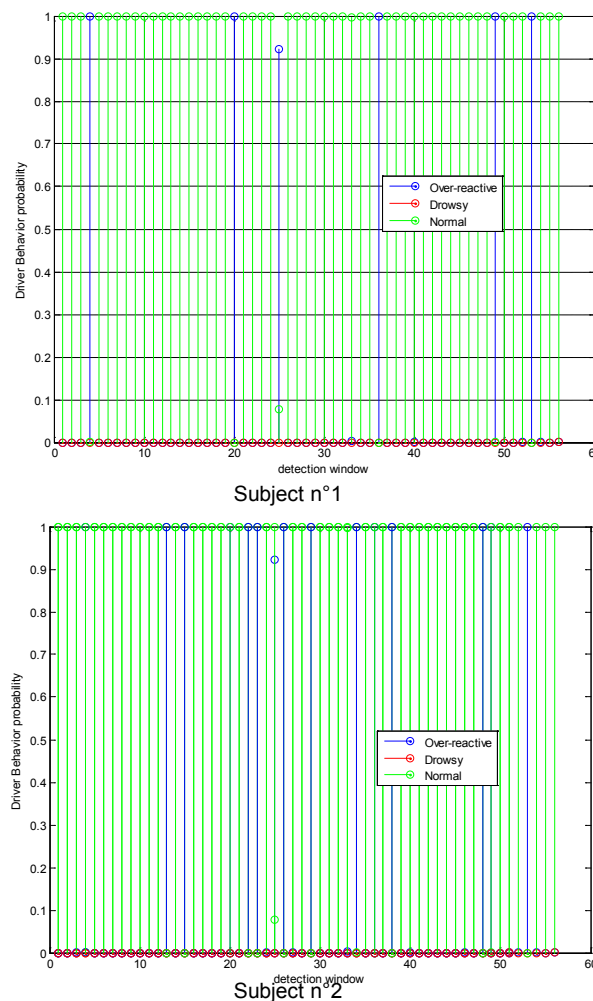


Figure 7. Behaviors detection by HMM

We note the detection of only two behaviors: the "over-reactive" and "normal". The non-detection of "drowsy" behavior seems logical, since the driving scenario takes only 15 min. The results show that the subject n° 2 has a more pronounced "over-reactive" behavior than the subject n°1.

The vertical eye movements of the two subjects n°1 and n°2 during a window's detection time are shown in "Fig.8". The pupil diameter evolution of the subjects is shown in "Fig.9".

The observed operator's blinks are immediate and their maximum number, detected during a window's time is 3. The human operator ocular activity is consid-

ered "small". This observation is instantiated in the BN with the behavior detection results.

The reaction time of the both drivers is measured and classified in "small" category; it is instantiated in the "driver reaction" node. The sensation seeking is considered "small" for the two subjects. After inference in the BN, the driver's state detection results are shown in "Fig.10".

The state of the subject n°1 is mainly estimated "normal" with a probability of $\sim 86\%$ during the driving scenario. The same subject presents the "over-reactivity" state on periods of 10 seconds at maximal; the probability of this state does not exceed the 60%. The apparition of these states is linked to the variation moments of the speed references. The driver does not present a real "drowsy" state, because the probability of this state never exceeds 10%.

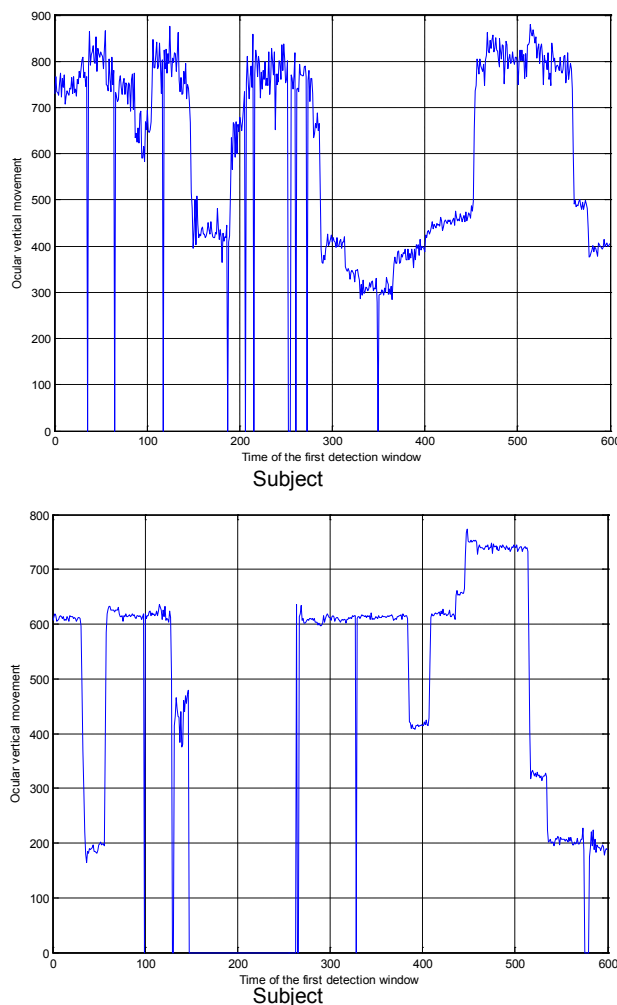


Figure 8. Vertical ocular movement evolution of drivers

The detection results using the proposed approach lead to the same observations on the subject n°2, but this operator has more states of "over-reactivity" that take longer time.

Remarks:

The non-detection of the drowsy state is realistic, as justified by the short time of the driving scenario that could not have a significant impact on the driver vigilance.

As the information distribution is bidirectional in a BN (Causes \leftrightarrow Effects). It is possible to deduce the ocular activity characteristics of an “over-reactive” operator. The preliminary results show that a driver having a "normal" driving behavior and being in an "over-reactive" state presents a "small" ocular activity and a "small" time reaction with respective probabilities of 0.54 and 0.85. This result can be analyzed to characterize the state of an ‘over-reactive’ operator.

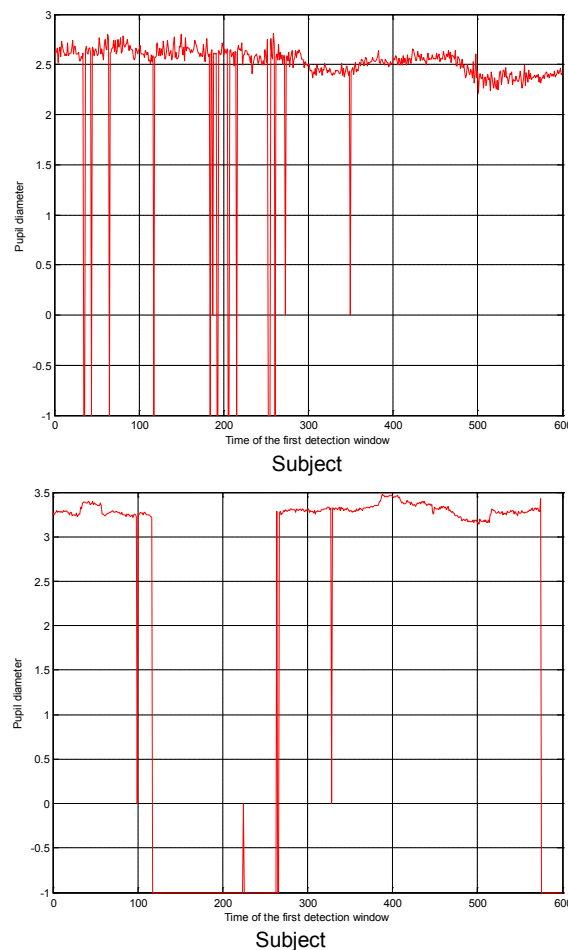


Figure 9. Pupil diameter evolution of drivers

Conclusion

In this article, an approach of drivers' state detection in rail systems is proposed. The detection methodology is made in two steps. The first one consists in the construction of the hidden Markov models for the identification of the driver's "typical" behaviors. The second one consists in the data fusion by a Bayesian network,

allowing a reliable detection despite the heterogeneity and the uncertainty of the measures.

The work presented in this article is a feasibility study of the proposed methodology, tested on data obtained during experiments on railroad simulation platform COR&GEST with encouraging results.

In the future works, the use of other endogenous or exogenous information sources such as the facial recognition system is envisaged. The objective is to improve the detection approach by multiplying the discrimination possibilities of the observed data. Besides, we plan to increase the tests number and request expert drivers.

The approach validation on another train driving simulator (with a mobile base of six degrees of freedom for the vibratory movements restitution "ILLUSIO" of LAMIH at Valenciennes is programmed.

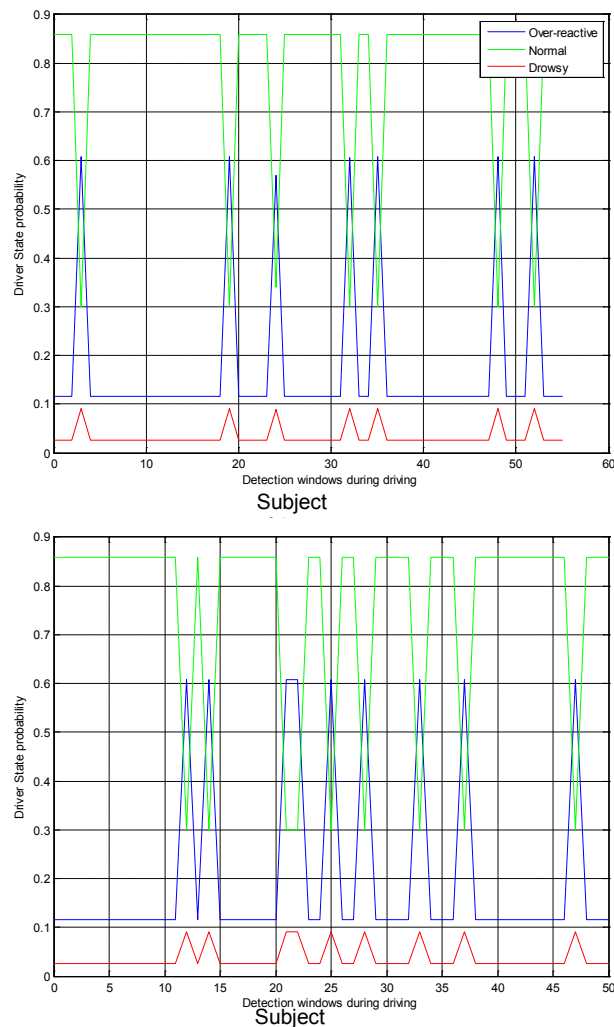


Figure 10. States detection by BN

References

- Berdjag D., Vanderhaegen F. Étude de la stabilité humaine dans un contexte ferroviaire, Acte de 18ème Congrès de Maitrise des Risques et Sureté de Fonctionnement, Tours, France, octobre 2012.
- Bonnefond A. et al. Behavioral reactivation and subjective assessment of the state of vigilance, *International Journal of Occupational Safety and Ergonomics*, vol. 12, n° 3, pp. 221-229, 2006.
- Chin-Teng L. et al. EEG-Based drowsiness estimation for safety driving using independent component analysis, *IEEE transactions on circuits and systems I*, vol. 52, n°12, pp. 2726-2738, 2005.
- Daza I. G. et al. Drowsiness monitoring based on driver and driving data fusion, 14th International IEEE Conference on Intelligent Transportation Systems, Washington, USA, 2011.
- Dong Y. et al. Driver inattention monitoring system for intelligent vehicles: a review, *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, n°2, pp.596 – 614, 2011.
- Foot R. et al. Questions raised on the design of the « dead-man » device installed on trams, *Cognition Technology and Work*, vol.10, n°1, pp.41-51, 2008.
- Gadepally V. et al. Driver/vehicle state estimation and detection, 14th International IEEE Conference on Intelligent Transportation Systems, Washington, USA, 2011.
- Igarashi K. et al. Biometric Identification Using Driving Behavioral Signals, *IEEE International Conference on Multimedia and Expo*, Taiwan, 2004.
- Krajewski J. et al. Steering wheel behavior based estimation of fatigue, 5th International Driving Symposium on Human Factors in Driver Assessment Training and Vehicle Design, Big Sky, Montana, USA, 2009.
- Kurt A. et al. Hybrid-State Driver/Vehicle Modeling Estimation and Prediction, 13th IEEE International Annual Conference on Intelligent Transportation Systems, Portugal, 2010.
- Larue G.S. et al. A model to predict hypovigilance during a monotonous task, *Proceedings of the Australasian Road Safety Research Policing and Education Conference*, Sydney, Australia, 2009.
- Liang Y. et al. Real-time detection of driver cognitive distraction using support vector machines, *IEEE Transactions On Intelligent Transportation Systems*, vol. 8, n°2, pp. 340-350, 2007.
- Naïm P. et al. Réseaux Bayésiens, Eyrolles, 2004.
- Oppenheim I. et al. Critical state of the art and unified models of driver behavior, Work package n°1, ITERATE European Project, 2010.

- McRuer D.T. , Krendel E.S. Mathematical models of human pilot behavior, Advisory Group for Aerospace research and development, AGARDograph n°188, Technical editing and reproduction ltd Hardford House, London, 1974.
- Qiang J. et al. A probabilistic framework for modeling and real-time monitoring human fatigue, IEEE Transactions on Systems Man and Cybernetics, vol.36, n°5, pp. 862-875, 2006.
- Qiang J. et al. Real-Time Eye, Gaze, and Face Pose Tracking for Monitoring Driver Vigilance, Real-Time Imaging, vol. 8, n°5, pp. 357–377, 2002.
- Rabiner L.R. A tutorial on hidden Markov models and selected applications in speech recognition, proceedings of the IEEE, vol.77, n°2, pp. 257-286, 1989.
- Rachedi N.D. et al. Détection de l'état d'un opérateur humain dans le contexte de la conduite ferroviaire, Acte de 18ème Congrès de Maitrise des Risques et Sureté de Fonctionnement, Tours, France, 2012.
- Richard P. et al. Vers le concept de stabilité humaine pour l'amélioration de la sécurité des transports, 17ème Congrès de Maîtrise des Risques et de Sûreté de Fonctionnement, La rochelle, France, 2010.
- Rigoll G. et al. A Comparison between Continuous and Discrete Density Hidden Markov Models for Cursive Handwriting Recognition, 13th IEEE Conference on Pattern Recognition, vol.2, pp.205-209, 1996.
- Schleicher R. et al. Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired?, Ergonomics, vol.51, n°7, pp. 982-1010, 2008.
- Szabolcsi R. Modeling of the human pilot time delay using Padé series, International Journal of Security Strategic Defense Studies and Military Technology, Vol. 6, n°3, pp.405-428, 2007.
- Vanderhaegen F. et al. A Benefit /Cost /Deficit (BCD) model for learning from human errors, Reliability Engineering & System Safety, vol. 96, n°7, pp.757-766, 2011.
- Verwey W. B. et al. Predicting drowsiness accidents from personal attributes, eye blinks and ongoing driving behavior, Elsevier, Personality and Individual Differences, vol.28, n°1, pp.123-142, 2000.
- S. Boverie, A. Giralt, Driver vigilance diagnostic based on eyelid movement observation, Proceedings of the 17th IFAC World Congress, Seoul, Korea, 2008.

Interactive tabletops as new supports for collective diagnosis?

Christophe Kolski¹, Yoann Lebrun^{1,2}, Bako Rajaonah¹, Sophie Lepreux¹, René Mandiau¹

Keywords: tabletop; tangible interaction; collective diagnosis; Collective Diagnosis Architecture (CDA), RFID

Abstract

The central question described in this paper concerns the possible use of interactive tabletops as new supports for collective diagnosis. A model called Collective Diagnosis Architecture (CDA) is proposed. An innovative application developed on two interactive tabletops equipped with RFID technology illustrates the proposition. Research ways are also envisaged.

Introduction

The design and evaluation of new interactive surfaces (such as tabletops), as well as the study of tangible interactions lead to new research ways, with dedicated conferences (TEI and ITS particularly), special issues in journals, books, and so on (Müller-Tomfelde, 2010). Interactive tabletops let consider new interactive applications usable in collective centralized and distributed situations. Two main configurations are possible: in the first one, several people interact on the same application around one tabletop; in the second one, through the use of connected tabletops, it is possible to interact with several people located in remote locations (Lepreux et al., 2011).

Knowing that diagnosis is a very vast and rich domain (see for instance Simpson & Sheppard, 1994), the question presented in this paper is the following: may interactive tabletops become new supports for collective diagnosis?

Section 2 recalls related work about diagnosis applications on tabletop. Section 3 proposes an architecture for collective diagnosis using one or more tabletops. Section 4 gives illustrations through a road traffic simulation. The last section concludes and proposes research perspectives.

Related work

Many applications on interactive tabletops are developed with educational motivations, but very few studies concerned by explicit or implicit diagnosis aspects are

¹ LAMIH-UMR CNRS 8201, University of Valenciennes, Le Mont Houy, 59313 Valenciennes cedex 9, France, {firstname.lastname}@univ-valenciennes.fr

² PRL-CCI Grand Hainaut, 3 avenue Sénateur Girard, BP 80577, 59308 Valenciennes Cedex, France
lebrun.yoann@gmail.com

currently described in the literature. Several researchers consider that tangible interaction can help to support diagnosis; for instance, with tangible paper interfaces, the tangible nature of paper can be used to extract information about the learning activity: (Bonnard et al., 2012) propose an augmented reality system for pupils in primary schools to explore the classification of quadrilaterals based on sheets, cards, and cardboard shapes. From the detected locations of the various user interface elements, they show how to extract indicators about problem solving, hesitation, difficulty levels of the exercises, and the division of labor among the groups of pupils.

In previous works, we have developed and tested in nursery school an application on TangiSense; this tabletop allows interacting with tangible objects. This interaction is particularly interesting to facilitate the interactions by children. The application allows children between 3 and 5 years with learning primary colors. This application proposes a set of statistics with time and location of tangible objects in order to support the teacher to diagnose potential problems (Kubicki et al., 2011).

Another application on interactive tabletop, named SimMed, allows medical students to collaboratively diagnose and treat a virtual patient. The hybrid user interface combines elements of virtual reality with multitouch input (von Zadow et al., 2013).

Tabletops let envisage new types of collective activities. Cognitive models are in consequence necessary for the analysis and the design of new applications on such interactive supports.

Proposition

The Collective Diagnosis Architecture (CDA) is an extension of the individual diagnosis and decision-making model (Hoc & Amalberti, 1995) to cope with collective diagnosis situations, especially in multi-agent systems. The core of the proposed architecture is the common knowledge base and the common work space. The common knowledge base is made up of the minimal common knowledge units (definitions, basic explanations, basic rules, etc.) required to define goals and to formulate hypotheses. The common work space (Pacaux-Lemoine & Debernard, 2002) is built up from facts and updated parameters, priorities, current states of the resources, and hypotheses related to the diagnosed situations. Interactive tabletops could be used to test hypotheses on current and anticipated situations and help to select the most appropriate actions with regard to the main goals.

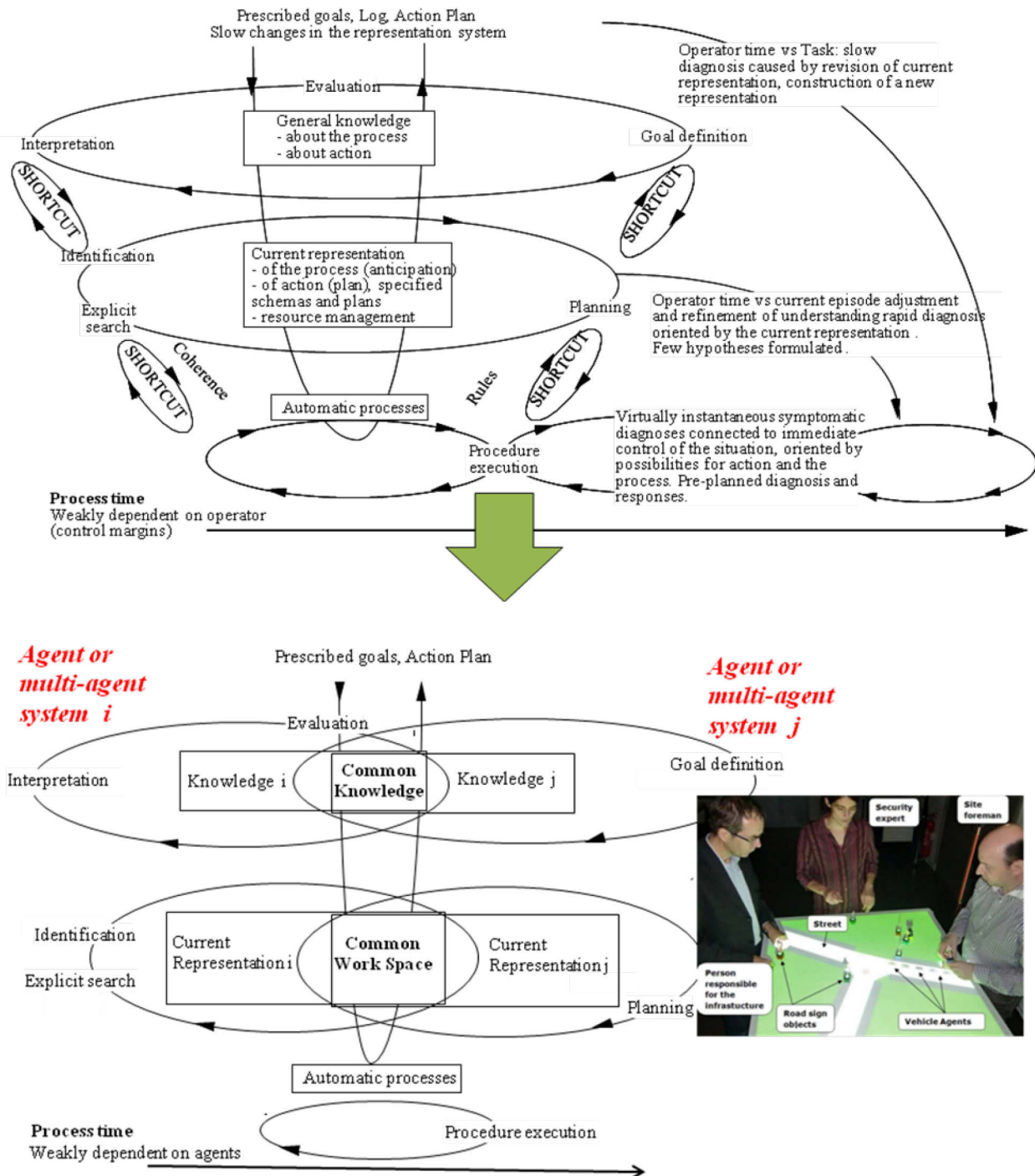


Figure 1: From the multi-level diagnosis architecture (Hoc & Amalberti, 1995, p. 96) to the Collective Diagnosis Architecture (CDA).

Case study

This section presents a case study related to the distributed simulation of road traffic management. We propose to take into consideration the architecture presented in the previous section to implement the simulator on two interactive tabletops, called TangiSense. Initially, the simulator has been designed on one TangiSense tabletop (Kubicki et al., 2013). This tabletop integrates RFID technology for the detection of objects equipped with tags. The tabletops are connected on a network

to communicate remotely. Figure 2 presents the road traffic simulator distributed on two tabletops with two user groups. The simulator uses the concepts of tangible and virtual agents (Lebrun et al., 2013) and aims at testing different hypotheses about the traffic: waiting time reduction in crossroads, accident situation management, infrastructure changes, and so on. Each tabletop is managed by a multi-agent system (MAS), which allows a common workspace (e.g. the entire simulator) and a current representation (e.g. a local representation for each tabletop).

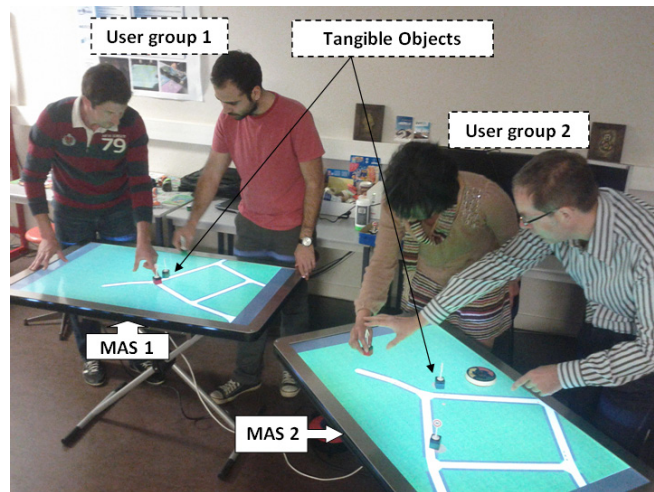


Figure 2: Road traffic management on two TangiSense tabletops

The common work space for the two user groups consists to collaborate on the same road map with the same initial traffic flows. These traffic flows are generated during the initialisation step of the application. The vehicles move randomly or according to a set of objectives. The current representation regarding a particular tabletop consists in selecting a road traffic area (e.g. road, crossroad) in order to modify the road signs. The user groups can collaborate on a common problem or on different spatial areas. The representation of the workspace is changed using tangible objects of interaction. These objects are used by being moved to enlarge the map, show the street name or speed for the different vehicles, etc. They are associated to a tangible agent and used to facilitate the interactions between the users and the table. The graphical representation for the different groups can be different even if the overall functioning of the simulator is identical in the sense that the vehicles associated with the agents have the same spatial coordinates in the MAS 1 and MAS 2.

Figure 3 is a graphical representation of the user group 1 (with respect to Figure 2) at a given moment of the traffic (i.e. the traffic). The action plan for this user group composed of two people is to reduce the traffic jam at a crossroad with tangible road signs objects. These tangible objects are represented in the figure by the traffic lights but they can also be characterized by a stop, a yield, etc.

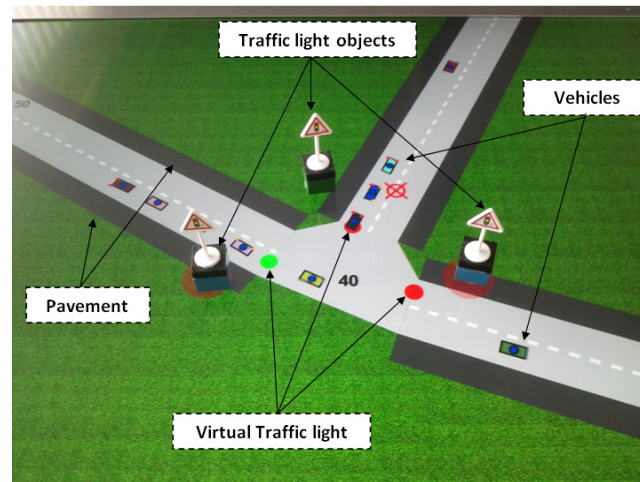


Figure 3: Current representation of the crossroad management

To be considered in the simulation, these objects should be placed on a pavement on each entry. When an object is detected by the support, a colored circle indicates that the system has taken the object into account. Every crossroad is associated to a virtual agent which manages all the entries with roles. This agent validates the user action when entrances are modified and executes the entrance behaviour. In the case of a crossroad with traffic lights, vehicles move then according to the lights state. During the simulation, the agent responsible for the crossroad analyzes data and provides statistics that allow the users to interpret particular situations. The statistics contain information about crossroad on entries and exit points of vehicles, the waiting time of a vehicle at an intersection, the maximum length of a lane and the average waiting time for each entry.

Conclusion

New interactive applications implemented on tabletops are more and more proposed and evaluated. Some of them allow users to interact with virtual and tangible objects in collective situations; such applications lead to new conceptual and use problems. In consequence the design and evaluation teams have to be supported by new cognitive models and architectures. With a view of collective diagnosis activity on tabletop(s), we have proposed in this paper the Collective Diagnosis Architecture (CDA), derived from the multi-level diagnosis architecture proposed by Hoc and Amalberti (1995). Illustrations through a road traffic simulation distributed on two tabletops equipped with RFID technology have been given. We think that tabletops may allow a better collaboration between different human experts for collective diagnosis problems. We consider that the collective diagnosis requires not only to produce shared knowledge by direct communications, but also to propose a simulation platform in which the effects of the experts' actions/decisions are directly visible (using tangible objects).

This paper is a starting point for an extensive work aiming at improving the Collective Diagnosis Architecture. Future studies and evaluations would be necessary in different application domains.

Acknowledgements

This research was partially financed by: Nord/Pas-de-Calais Region, CNRS, FEDER, CISIT, GDR E Hamasyt, ANR (IMAGIT project, financial IMAGIT support: ANR-10-CORD-017). The authors thank also Emmanuel Adam for his strong contribution to the simulator illustrating the proposition.

References

- Bonnard, Q., Jermann, P., Legge, A., Kaplan, F., & Dillenbourg, P. (2012). Tangible paper interfaces: interpreting pupils' manipulations. *Proc. Int. conf. on Interactive tabletops and surfaces*, ACM, New York, 133-142.
- Hoc, J.M., & Amalberti R. (1995). Diagnosis: some theoretical questions raised by applied research. *Current Psych. of Cogn.*, 14(1), 73-101.
- Kubicki, S., Lepreux, S., & Kolski, C. (2011). Evaluation of an interactive table with tangible objects: Application with children in a classroom. *Proceedings 2nd Workshop on Child Computer Interaction "UI Technologies and Educational Pedagogy"*, at CHI'2011, Vancouver, may.
- Kubicki, S., Lebrun, Y., Lepreux, S., Adam, E., Kolski, C., Mandiau, R. (2013). Simulation in Contexts Involving an Interactive Table and Tangible Objects. *Simulation Modelling Practice and Theory*, 31, 116–131.
- Lebrun, Y., Adam, E., Mandiau, R., Kolski, C. (2013). Interaction between tangible and virtual agents on interactive tables: Principles and case study. In *ANT 2013, 4th Int. Conf. on Ambient Systems, Networks and Technologies* (June 25-28, Halifax, Canada), Elsevier, 32-39.
- Lepreux, S., Kubicki, S., Kolski, C., & Caelen, J. (2011). Distributed Interactive Surfaces: A step towards the distribution of tangible and virtual objects. In *Distributed User Interfaces, Designing Interfaces for the Distributed Ecosystem*, Springer, 133-143, ISBN 978-1-4471-2270-8.
- Müller-Tomfelde, C. (2010). *Tabletops; Horizontal Inter. Displays*. Springer.
- Pacaux-Lemoine M.P., & Debernard S. (2002). A Common Work Space to support the Air Traffic Control. *Control Eng. Practice*, 10(5), 571-576.
- Simpson, W.R., & Sheppard, J.W. (1994). *System Test and Diagnosis*. Kluwer Academic Publishers, Norwell, Massachusetts.
- Von Zadow, U., Buron, S., Harms, T., Behringer, F., Sostmann, K., & Dachsel, R. (2013). SimMed: combining simulation and interactive tabletops for medical education. In *Proc. CHI '13*, ACM, NY, 1469-1478.

Toward the reverse comic strip concept to diagnose human state based on sound and facial recognition systems

Frédéric Vanderhaegen¹

Keywords: recognition systems, emotion image, sound image, human state diagnosis, reverse comic strip

Abstract

The paper presents the reverse comic strip as an original way to diagnose human states. The reverse comic strip is produced a posteriori and is composed by sound and emotion images that relate to the human state. The sound images are built by a system that detects the evolution of the microphone diaphragm deformation produced by the human voice. The emotion images are selected manually among a list of predefined images. A case study illustrates the feasibility of such an approach. It consists in interpreting the meaning of some information of a rail signalling system. Two main human states can then be diagnosed by using the reverse comic strip: the lack or doubt of knowledge, and the perseveration behaviour.

Introduction

The comic strip relates to narrative sequences based on texts and pictures or images. It is used for specific purposes such as education, communication or research (Lawson, 2006; Tatalovic, 2009). The reverse comic strip consists in building an a posteriori comic strip based on the on-line human behaviours (Vanderhaegen, 2013). The building of such a comic strip takes into account data, for instance, from facial and sound recognition systems. A facial recognition system leads to create an emotion image and a voice recognition system leads to capture a sound image.

The emotion images display different kinds of emotional characteristics such as frightened or scared, worried or furious, surprised or astonished, or puzzled or dissatisfied. They can be generated automatically by using supports such as facial recognition systems (Legrand et al., 2012). The sound image can come from difference sources. For instance, it can be built from specific sounds of the voice, and each sound can be identified into a particular word or sound database. The sound image can also be associated to a particular evolution of the voice frequencies.

¹ LAMIH-UMR CNRS 8201, University of Valenciennes, Le Mont Houy, 59313 Valenciennes cedex 9, France, frederic.vanderhaegen@univ-valenciennes.fr

The emotion images presented on this paper are selected manually and the sound image is based on the evolution of the microphone diaphragm vibrations presented on (La Delfa, 2013).

Both emotion and sound images are used as a dynamic reverse comic strip to support the diagnosis of the knowledge state of human operators. A feasibility study is then proposed with a practical example related to the possible interpretations of railway signalling pictures. It is treated in terms of emotion and sound to identify possible lack of knowledge, doubt on knowledge or erroneous application of knowledge. The paper is divided into four parts. The first part presents the reverse comic strip concept. The second part details the case study. The last two parts discuss about the possible interpretation of the reverse comic strip content in terms of diagnosis of human states. Two states are identified: the lack or doubt of knowledge, and the perseveration behaviour.

The reverse comic strip principle

The reverse comic strip concept (Vanderhaegen 2013) is a nice example of tool to analyse human behaviours. It consists in building a comic strip after capturing real-time data from human behaviours.

For instance, these data can relate to the facial and the voice characteristics, Figure 1.

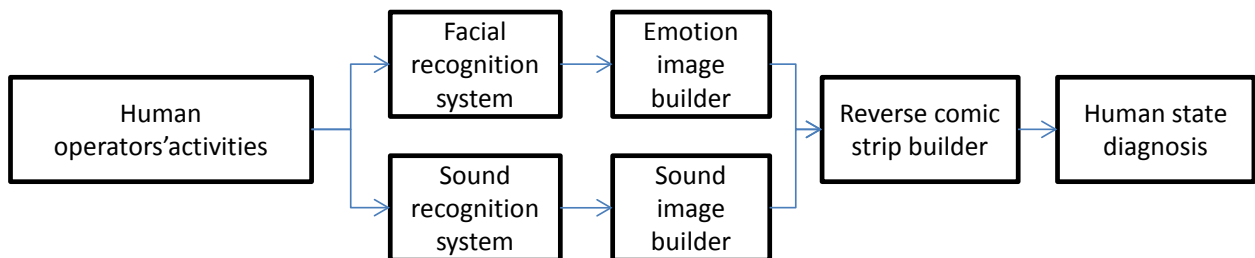


Figure 1: An example of a reverse comic strip builder to diagnose human state.

Both facial and sound recognition systems are then required in order to build the emotion images and the sound variation images. Figure 2 proposes some examples of such emotion images and sound variation images.

The facial recognition system recognizes face pictures of the reverse comic strip. It is associated to specific emotions such as frightened or scared, worried or furious, neutral or undetermined, etc.

The sound recognition system recognizes the sound variation of the comic strip based on the sounds produced by the human voice. It can be associated to a particular word or sound database, to a particular evolution of voice frequencies, or the evolution of microphone diaphragm vibrations.

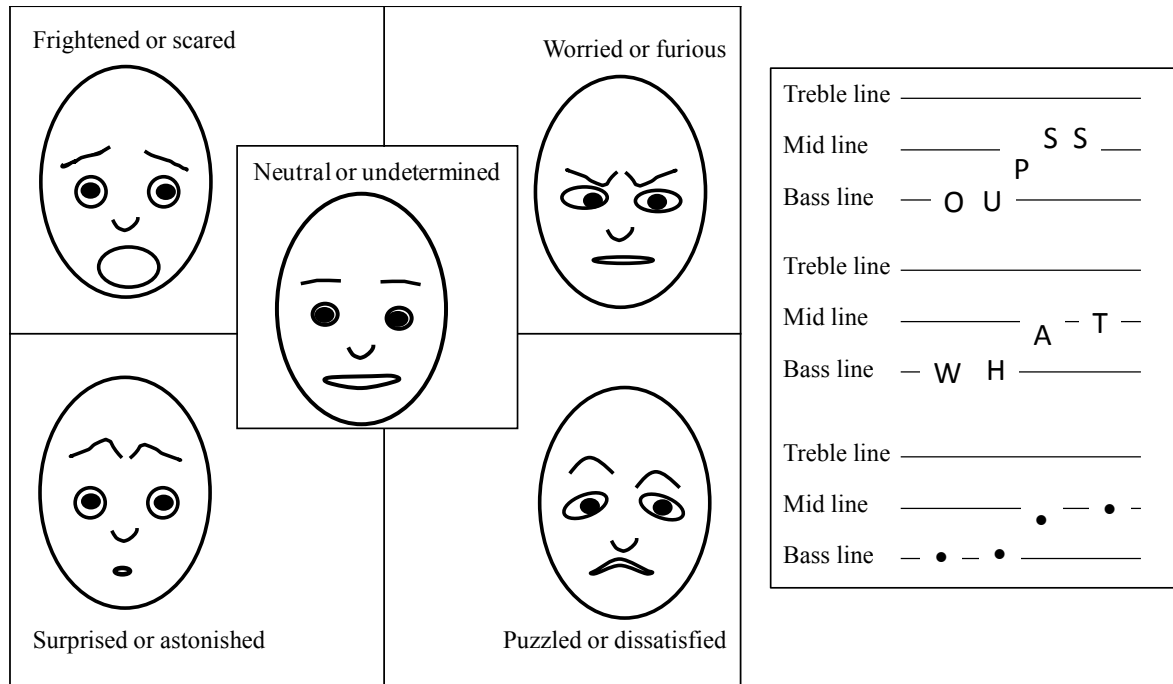


Figure 2: Examples of emotion and sound variation images.

This paper applies the reverse comic strip concept by using 1) the variation of the microphone diaphragm produced by the human voice and 2) the manual selection of the emotion images made by the subjects.

A case study to diagnose human state

The case study consists in displaying three railway based pictures during 5 seconds and to try to explain their meaning by producing knowledge. Knowledge was associated to a rule such as $A \rightarrow B$, i.e. if A then B.

The pictures of this case study are presented on Figure 3. Three subjects participated to this experimental protocol that consists in determining the meaning of the lights burning, i.e. when the lights are on, what does it mean?

The vibration of the microphone diaphragm was recorded automatically by using the application developed on (La Delfa, 2013). In parallel, the subjects were invited to select the emotion images associated to their states during the research of the meaning of the burning lamps.

After the experiment, the subjects explained what happened during this research process.



Figure 3. The railway based pictures of the experimental protocol.

Two kinds of diagnosis were done a posteriori: the diagnosis of a lack or doubt of knowledge and a diagnosis of perseveration behaviour.

Diagnosis of a lack or doubt of knowledge

Two subjects had similar behaviour and explanation during this exercise. Their corresponding reverse comic strips are presented on Figure 4.

Firstly, facing picture 1 (i.e., display 1), due to a lack of knowledge, they tried to discover or to recognise the meaning of the lighted lamps and then they produce a rule: if the lights are on, then the train will stop at the corresponding train stations. The variation of the sound and emotion images relate to this occurrence of a lack of knowledge and to this research process. The end of the process is determined by a stability of the sound image and by the neutral or undetermined emotion image.

Facing picture 2 (i.e., display 2), after having a doubt about the meaning of the lighted lamps, they confirmed that it is the same rule. The doubt of this research process appears with the worried or the surprised emotion images and with the production of sounds by the subjects.

For picture 3 (i.e., display 3), they initially thought it was the same rule, and after detecting their error due to another lack of knowledge, they produced a new rule: if the lights are on, then they indicate where the trains are located on the line. Both subjects selected the neutral or undetermined emotion image at the beginning of the research process because they were sure it was the same rule. The detection of their error of interpretation and the production of a new research process is indicated by a variation of the microphone diaphragm deformation and by the worried or

puzzled emotion images. The end of this process occurs when the sound image became stable and when the subject selected the neutral or undetermined emotion image.



Figure 4. The reverse comic strip of the subjects 1 and 2.

Diagnosis of perseveration behaviour

The last subject was initially astonished before recognizing the meaning of the burning lamps after the display of the picture 1. This was expressed verbally during the experiment.

After this recognition process, the sound becomes stable and the selected emotion image is neutral or undetermined. The produced rule was: if the lamps are on then the train will stop at the stations indicated on the signalling panel.

This rule was confirmed after the display of the pictures 2 and 3. Despite a doubt illustrated by a variation of the microphone diaphragm deformation and by the astonished emotion image, the sound image becomes stable again and the emotion image is neutral or undetermined.

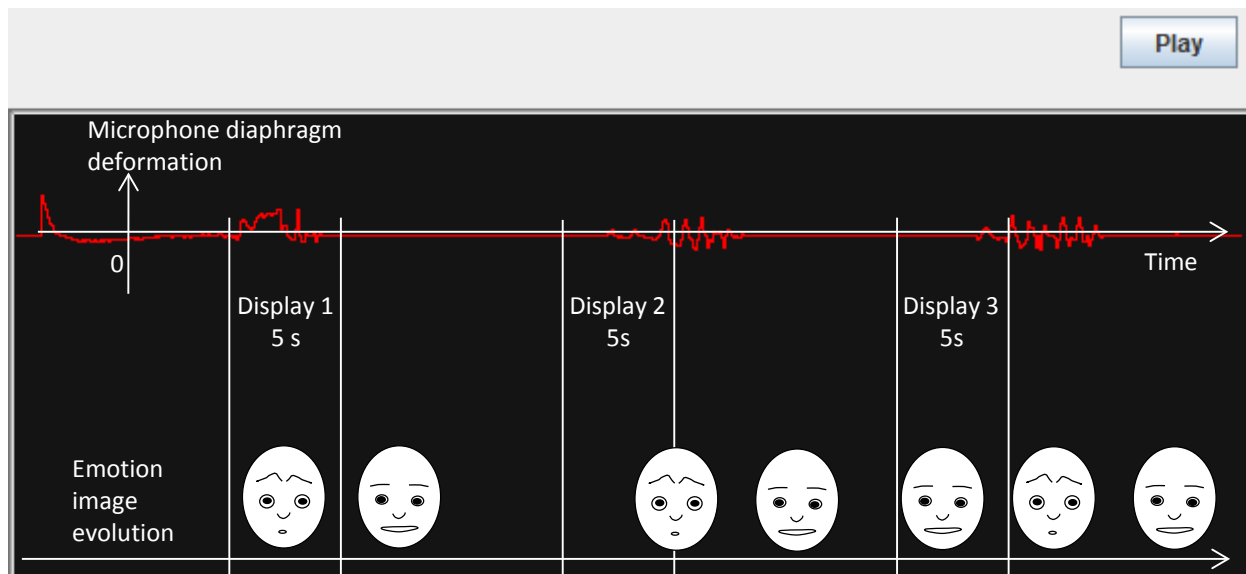


Figure 5. The reverse comic strip of the subject 3.

After the last display, the subject did not produce any new rule and confirmed the explanation of the burning lamps by the first rule defined after the display 1. This behaviour is similar to the so-called perseveration behaviour described on (Dehais et al., 2012). It relates to the erroneous application of knowledge.

Conclusion

This paper presented an original way to diagnose human states based on facial and sound recognition systems. This is so-called the reverse comic strip because it is built on a posteriori process displaying both emotion and sound images in the course of the human activities. A case study illustrated the feasibility of such a diagnosis approach based on the reverse comic strip. A stability of the sound image based on the microphone diaphragm deformation is associated to the neutral or undetermined emotion image. A high level of deformation of the microphone diaphragm is linked to puzzled, surprised, worried or astonished emotion images. A lack of knowledge or a doubt of knowledge, and the perseveration behaviour were then diagnosed by using this reverse comic strip based on sound and emotion images. An extended experimental protocol will be organised in order to produce automatically the emotion images and to validate more deeply the efficiency of the reverse comic strip for the real-time human state diagnosis.

Future researches will translate the reverse comic strip concept in terms of stability (Zieba et al., 2011; Richard et al. 2010; Vanderhaegen et al. 2012). Stability of

emotion images and/or sound images will be an interesting prospective study to identify the relationship between the stability of emotion or sound images and the diagnosis of human states. Thresholds will be defined in order to identify an out of stability or an overstability and to diagnose human states automatically. Cooperation and learning processes described on (Vanderhaegen, 1997; Vanderhaegen, 1999; Vanderhaegen et al., 2009; Vanderhaegen, 2012) may also be useful to improve the autonomy of human-machine systems or to facilitate the diagnosing their state.

Acknowledgements

The present research work has been supported by European Research Network HAMASYT (Human-Machine Systems in Transportation and Industry), and by the French state through the Agence Nationale de la Recherche as part of the Investissement d'Avenir program number ANR-11-IDFI-0037. The author gratefully acknowledges the support of these institutions.

References

- Dehais, F., Causse, M., Vachon, F. & Tremblay, S. (2012). Cognitive conflict in human-automation interactions: a psychophysiological study. *Applied Ergonomics*, 43(3), 588-595.
- La Delfa, S. (2013). Sound recognition system. Master Science research project, University of Valenciennes.
- Lawson, R. A. (2006). Teaching economic principles with comic strips. *Journal of Private Enterprise*, 22(1), 168-176.
- Legrand, C., Richard, P. Benard, V., Vanderhaegen F. & Caulier, P. (2012). Diagnosis of human operator behaviour in case of train driving: interest of facial recognition. *Proceedings of the 30th European Annual Conference on Human Decision Making and Manual Control (EAM2012)*. Braunschweig, Germany, September.
- Richard, P., Benard, V., Caulier, P. & Vanderhaegen, F. (2010). Toward the "Human Stability" in transportation domain: concepts and objectives. *Proceedings of the 11th IFAC/IFIP/IFORS/IEA symposium on Analysis, Design and Evaluation of Human-Machine Systems*, August-September 2010, Valenciennes, France.
- Tatalovic, M. (2009). Science comics as tools for science education and communication: a brief, exploratory study. *Journal of Science Communication*, 8(4), 1-17.
- Vanderhaegen, F. (1997). Multilevel organization design: the case of the air traffic control. *Control Engineering Practice*, 5(3), 391-399.

- Vanderhaegen, F. (1999). Multilevel allocation modes - Allocator control policies to share tasks between human and computer. *System Analysis Modelling Simulation*, 35, pp. 191-213.
- Vanderhaegen, F. (2013). Toward a reverse comic strip based approach to analyse human knowledge. *Proceedings of the 12th IFAC/IFIP/IFORS/IEA symposium on Analysis, Design, and Evaluation of Human-Machine Systems*, August 11-15 2013, Las Vegas, USA.
- Vanderhaegen, F. (2012). Cooperation and learning to increase the autonomy of ADAS. *Cognition, Technology & Work*, 14, 61-69.
- Vanderhaegen, F., Caulier, P. (2011). A multi-viewpoint system to support abductive reasoning. *Information Sciences* 181(24), 5349-5363.
- Vanderhaegen, F., Polet, P., Zieba, S. (2009). A reinforced iterative formalism to learn from human errors and uncertainty. *Engineering Applications of Artificial Intelligence* 22(4-5), 654-659.
- Vanderhaegen, F., Zieba, S., Enjalbert, S., Polet, P. (2011). A Benefit/Cost/Deficit (BCD) model for learning from human errors. *Reliability Engineering & System Safety* 96(7), 757-76.
- Zieba, S., Polet, P., Vanderhaegen, F. (2011). Using adjustable autonomy and human-machine cooperation for the resilience of a human-machine system – Application to a ground robotic system. *Information Sciences*, 181, pp. 379-397.

Wie viel ist zu viel? Empfehlungen für die Anzahl von Bedienelementen auf kleinen Displays

Nils Backhaus und Stefan Brandenburg¹

Schlüsselwörter: Displaygestaltung, Graphical User Interfaces, Kognitionspsychologische Grundlagen und Anwendung, ökologische Anzeigengestaltung

Zusammenfassung

Die Zunahme der Nutzung mobiler, kleiner, digitaler Endgeräte stellt Displaydesigner vor die Herausforderung, auf kleinstem Raum eine Vielzahl von Funktionen zur Verfügung zu stellen. Befunde kognitionspsychologischer Grundlagenforschung und des ökologischen Interface-Designs (ecological interface design, EID) werden genutzt, um zu überprüfen, wie sich die Anzahl von Bedienelementen auf die wahrgenommene Gebrauchstauglichkeit (Usability) auswirkt. Klassische kognitionspsychologische Befunde (bspw. Sternberg, 1969) unterstützen die Hypothese "less is more" - je geringer die Anzahl an Bedienoptionen, desto besser ist die Leistung. Dagegen postuliert das EID eher "adequate is better". D. h., dass auch zunächst überflüssig erscheinende Systemelemente die Gebrauchstauglichkeit positiv beeinflussen können, wenn dadurch die Bildung eines adäquaten mentalen Modells auf Seiten des Nutzers gesichert wird (z. B. Flach et al., 1998). Beide Hypothesen wurden in einem Experiment gegenübergestellt. 53 Probanden interagierten mit einer von drei unterschiedlich gestalteten graphischen Nutzerschnittstellen (graphical user interface, GUI) eines MP3-Spielers. Die Ergebnisse zeigen, dass der Mp3-Spieler am besten beurteilt wurde, der zu jeder Zeit alle für das korrekte mentale Modell benötigten Bedienelemente anzeigte. Demnach stützen die Resultate eher die Hypothese des EID. Da in der aktuellen Arbeit lediglich subjektive Daten erhoben wurden, sollte sich die weitere Forschung sowohl weiteren subjektiven als auch objektiven Messwerten widmen.

Einleitung

Mobile, digitale Endgeräte (z. B. Smartphones, Tablets) sind zunehmend Teil unseres Alltags. Das US-Marktforschungsunternehmen Gartner schätzt, dass in diesem Jahr 1,2 Milliarden mobile Endgeräte verkauft werden – die Anzahl vorhandener Geräte wird sich bis 2016 sogar verdreifachen (Gartner, 2012). Die Geräte müssen flach und leicht sein, die dementsprechend kleinen Displays dieser Geräte stellen hohe Herausforderungen an die Gestaltung von GUIs (Chittaro, 2006). Beispielsweise ist es schwierig zu entscheiden, wie viele Bedienelemente auf kleinstem Raum angezeigt werden sollten bzw. wie viele Bedienelemente vom Nutzer für eine erfolgreiche Interaktion benötigt werden (Mauney & Masterton, 2008). Ana-

¹ Technische Universität Berlin, Fachgebiet Kognitionspsychologie und Kognitive Ergonomie

log zu Long (1989) gehen die Autoren des vorliegenden Beitrages davon aus, dass Befunde aus vorherigen Arbeiten zur Bestimmung der optimalen Anzahl von Interaktionselementen für geringe Displaygrößen herangezogen werden können. Insbesondere die Ergebnisse zweier Perspektiven sind von Bedeutung, die der Grundlagenforschung im Bereich kognitionspsychologischer Informationsverarbeitung und die des EID.

"Less is more" – Befunde klassischer Experimente zur Informationsverarbeitung

Im Rahmen der kognitionspsychologischen Perspektive liefern klassische Reaktionszeitexperimente zu Informationsverarbeitungsprozessen (bspw. Sternberg, 1969) wichtige Hinweise zur Anzahl der präsentierten Interaktionselemente. Im Rahmen dieser Experimente sollten sich Probanden ein bestimmtes Stimulusset einprägen (Präsentationsdauer ca. 1 s). Im Anschluss wird ein einzelner Stimulus gezeigt. Der Proband soll möglichst schnell angeben, ob dieser im zu erinnernden Stimulusset enthalten war oder nicht. Die Reaktionszeit steigt dabei in Abhängigkeit der Größe des Stimulussets linear an (ausgehend von einem Basislevel für die Enkodierung und Ausführung der Reaktion). Dieser Befund zeigt, dass eine bestimmte Zeit zum Enkodieren der Informationen und zum Vergleich eines Zielstimulus mit einem Stimulusset notwendig ist. Jede zusätzliche Information erhöht diese kognitive Verarbeitungszeit. Dass Reaktionszeiten von wenigen Sekunden, wie bei Sternberg, auch für die Interaktion von Menschen mit mobilen Endgeräten von Bedeutung sind, zeigten bereits Arbeiten von Oulasvirta et al. (2005) oder auch Drewitz und Brandenburg (2010). Während Oulasvirta et al. (2005) ermittelten, dass Nutzer sich in verrauschten Umgebungen (bspw. lebhafter Straße) ca. 1–8 s mit ihrem mobilen Endgerät beschäftigten, um Informationen zu finden, stellten Drewitz und Brandenburg (2010) den Link von kurzen zeitlichen Verzögerungen (1-3 Sek.) bei der Interaktion mit einem mobilen Endgerät zu einem verändertem Nutzererleben her. Demnach legen die Befunde zur Sternberg-Aufgabe die Strategie „Weniger ist besser“ nahe, wenn es um die Anzahl von Bedienelementen auf einem mobilen Endgerät geht. Dementsprechend sollte nur das, was für die Interaktion unbedingt notwendig ist, angezeigt werden. Alle überflüssigen Bedienelemente würden die Suche nach dem richtigen Bedienelement verlängern und die Interaktion verlangsamen (vgl. Sternberg, 1969) und dadurch unmittelbares, interaktives Verhalten verhindern (Gray, 2006). Dieser Ansatz entspricht auch praxisnahen Empfehlungen zur Gestaltung von small-screen GUIs: "Keep the navigation narrow and shallow" (Mauney & Masterton, 2008, S. 344).

"Adequate is better" – Ökologische Anzeigengestaltung (EID)

Das EID entwickelte sich ursprünglich als Rahmenmodell für die Gestaltung von Interfaces komplexer Mensch-Maschine-Systeme. Dem Modell liegen zwei Ziele zugrunde: Zum einen darf die Beanspruchung aufgrund der Aufgaben nicht unnö-

tig durch das Interface erhöht werden. Zum anderen sollte die Displaygestaltung die Aufgabe und damit die kognitive Leistungsfähigkeit unterstützen (Vicente & Rasmussen, 1992). Im Gegensatz zur kognitionspsychologischen Perspektive würden die Prinzipien des EID nahelegen, mehr als das Minimum an Bedienelementen zu zeigen (bspw. Flach et al., 1998). Ziel ist es hierbei immer diejenigen Bedienelemente zu zeigen, die auch im mentalen Modell des Nutzers benötigt werden, um zu verstehen welche Prozesse im technischen System ablaufen und welche Handlungsmöglichkeiten er hat. Vicente und Rasmussen (1992) fassen die Idee des EID wie folgt zusammen: "Thus, an EID interface should not contribute to the difficulty of the task, and [...] it should support the entire range of activities that operators will be faced with" (S. 589). Obwohl EID zumeist im Bereich komplexer, dynamischer Prozesssysteme genutzt wurde (Wickens et al. 2012), passt der Ansatz auch zum GUI-Design mobiler Endgeräte (z. B. Lindgren et al., 2009).

Experimenteller Zugang

Stichprobe, Material und Durchführung

In diesem Zusammenhang wurde ein erster Versuch unternommen, um die Annahmen der beiden Perspektiven experimentell gegenüberzustellen. 53 Probanden (62% männlich, Alter: $M = 24.39$, $SD = 5.67$) wurde jeweils eine von drei Versionen eines MP3-Players vorgelegt, die sich nur in der Anzahl der Bedienelemente unterschieden (Abb. 1). Die erste Variante enthielt weniger als die benötigten Bedienelemente. Hier war (analog zur adaptiven Anzeigengestaltung, vgl. Benyon & Murray, 1988) der Wiedergabeknopf bei der Nicht-Wiedergabe und der Pause/Stopp-Button bei der Wiedergabe im GUI sichtbar (Abb. 1a). Die zweite Version des MP3-Spielers enthielt alle Bedienelemente, die für die Erledigung der drei Aufgaben (Lied anspielen, stoppen und weiterstellen) von Bedeutung waren (Abb. 1b). Die dritte Version enthielt zusätzliche, für die Aufgabenerledigung nicht benötigte Bedienelemente (Lautstärkereger, Abb. 1c). Die MP3-Player wurden auf einem iPAQ 214 mit einer Bildschirmgröße von 4" und einer Auflösung von 640x480 Pixeln präsentiert. Die Bedienung erfolgte über einen Stylus.

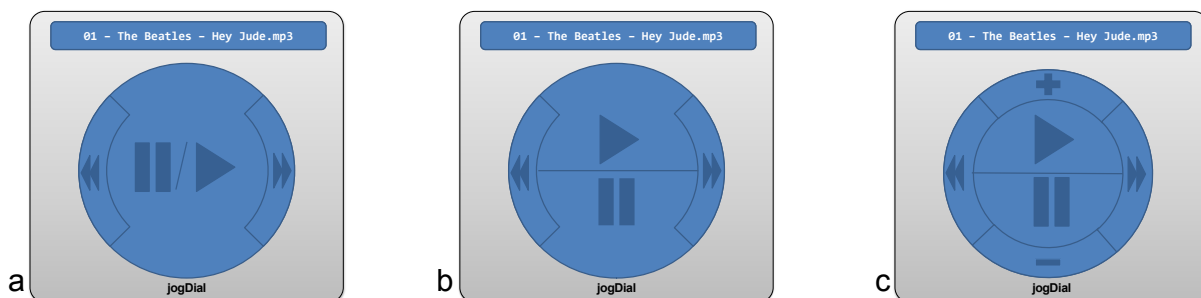


Abbildung 1: MP3-Player-Versionen, a) minimale Anzahl an Bedienelementen (min.) b) für Aufgaben erforderliche Anzahl an Bedienelementen (med.) c) mehr als erforderliche Anzahl an Bedienelementen (max.)

Um das Nutzererleben zu erfassen wurde der AttrakDiff verwendet (Hassenzahl, 2003). Der Fragebogen misst das Nutzererleben über 23 Items des semantischen Differentials auf vier Dimensionen (pragmatische Qualität, hedonische Qualität mit Bezug auf Identität und hedonische Qualität, mit Schwerpunkt Identität, Stimulation bzw. Attraktivität). Dabei betont die Skala der pragmatischen Qualität klassische Usability-Kriterien, wobei die hedonischen Qualitäten primär das "look and feel" erfassen sollen. Zur Erfassung der visuellen Ästhetik diente der VisAWI-S (Moshagen & Thielsch, 2011). Die Kurzskala des VisAWI ermöglicht die Erfassung der Wirkung des Layouts und der Gestaltung von Interfaces anhand von vier siebenstufig likert-skalierten Items.

Zu Beginn des Versuchs bekamen die Probanden eine kurze Einführung zur Benutzung des Mp3-Spielers. Die Aufgabenbeschreibung wurde ebenfalls auf dem Display des Geräts präsentiert. Die Versuchspersonen sollten typische Bedienungsaufgaben mit einem MP3-Player ausführen. Nach einem kurzen Anspielen (2–3 s) eines Songs (Play) sollten sie die Wiedergabe stoppen (Pause/Stopp) und zum nächsten Song weiterstellen (Forward) und diesen zweiten Song wieder anspielen (2–3 s) und stoppen. Anschließend füllten die Probanden die Fragebögen aus. Die Durchführung des Versuchs dauerte ca. 10 min.

Ergebnisse

Zur Analyse der Daten wurden einfaktorielle Varianzanalysen gerechnet (dreifachgestufter Faktor: Displayversionen). Abhängige Maße waren die unterschiedlichen Subskalen des AttrakDiff und des VisAWI-S. Ein(e) Teilnehmer(in) wurde aufgrund extremer Werte aus dem Datensatz entfernt. Die Ergebnisse (Abb. 2) zeigen, dass der MP3-Spieler mit der ausreichenden Anzahl von Bedienelementen am besten beurteilt wurde (AttrakDiff, pragmatische Qualität; $F(1,50) = 4.04$, $p = 0.04$, $f = 0.40$). In der Display-Komplexität (VisAWI) und dem Gefallen (AttrakDiff, hedonische Qualität) zeigten sich keine Unterschiede ($p_s > 0.05$; $F_s < 1$).

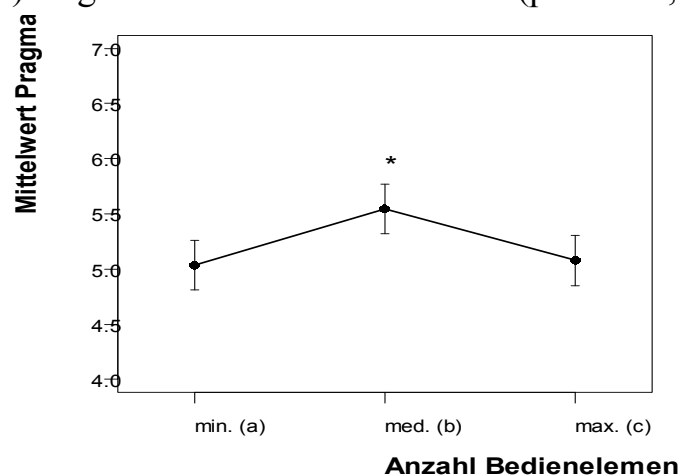


Abbildung 2: Mittelwerte der Bewertungen der drei MP3-Player-Versionen in Bezug auf pragmatische Qualität (+/- 1 Standardfehler), (*: $p < 0.05$)

Diskussion

Die vorliegende Untersuchung hatte zum Ziel zwei gegensätzliche Perspektiven auf das Interface Design von kleinen Screens gegenüberzustellen, die der kognitiven Ergonomie und die des EID. Zumindest auf der subjektiven Ebene wird deutlich, dass insbesondere der Ansatz des EID vielversprechend für die Gestaltung von Dialogen auf kleinen Bildschirmen scheint. Damit reiht sich das vorliegende Ergebnis in die Befunde früherer Arbeiten ein (z. B. Seppelt & Lee, 2007; Young & Birrel, 2012). So haben Seppelt und Lee (2007) beispielsweise aufgezeigt, dass die Perspektive des EID genutzt werden konnte, um ein Fahrerinformationssystem zu gestalten, das die Arbeitsweise des Adaptive Cruise Control (ACC) eindeutig an die Fahrer vermittelt. Interessanterweise wirken sich die Unterschiede in der Gestaltung nur auf die pragmatische Qualität des Interfaces aus - visuelle Ästhetik und hedonische Qualitäten sind hiervon unbeeinflusst. Es zeigt sich, dass die unterschiedliche Anordnung der Tasten nur auf die wahrgenommene Usability, aber nicht auf die ästhetische / hedonische Erfahrung auswirkte.

Zum derzeitigen Stand der Arbeiten ist allerdings weitere Forschung nötig, etwa um die Übertragbarkeit der Befunde auf komplexere Anzeigen zu überprüfen und sie mit objektiven Maßen zu validieren (z. B. Aufgabebearbeitungszeiten). Weiterhin bot das Interaktionsmedium in der aktuellen Arbeit, ein PDA mit Stylus, zwar umfangreiche Funktionalitäten für die Interaktion mit den GUIs, war aber im Vergleich zur weiter verbreiteten Touch-Bedienung mit Fingern eher nicht Stand der Technik. Außerdem war die Tiefe und Dauer der Interaktion mit den MP3-Player-Modellen relativ gering, wenn auch im Rahmen des Üblichen (vgl. Oulasvirta et al., 2005). Erste Hinweise bezüglich der Empfehlungen für die Gestaltung von kleinen Displays ergeben sich allerdings bereits aus der vorliegenden Studie. Beispielsweise reicht es nicht, nur die momentan benötigten Informationen im GUI aufzuzeigen. Vielmehr sollte nach der Devise verfahren werden: Zeige so viel wie möglich, allerdings nicht mehr als nötig. Wie viel das ist, lässt sich ebenfalls mit dem Ansatz des EID herausfinden. Methoden wie die Abstraction Hierachy Analysis (bspw. Stoner et al., 2003) weisen hier auf derartige Möglichkeiten hin.

Literatur

- Benyon, D., & Murray, D. (1988). Experience with Adaptive Interfaces. *The Computer Journal* 31(5), 465–473.
- Chittaro, L. (2006) Visualizing information on mobile devices. *IEEE Computer* 39(3), 40–45.
- Drewitz, U., & Brandenburg, S. (2010). From design to experience: Towards a process model of user experience. In: Lin, J.C., Lin, D.M. & Chen, H. (Eds.) *Proceedings of 20th Ppcoe*, Kaohsiung, Taiwan.

- Flach, J., Vincente, K., Tanabe, F., Monta, K., & Rasmussen, J. (1998). An ecological approach to interface design. *Proceeding of the HFES 1998*, 295–299.
- Gartner (2012). Gartner Says 821 Million Smart Devices Will Be Purchased Worldwide in 2012; Sales to Rise to 1.2 Billion in 2013 (Press release). Available online at: <http://www.gartner.com/newsroom/id/2227215>.
- Gray, W. D., Sims, C. R., Wai-Tat, F., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review* 113(3), 461–482.
- Hassenzahl, M. (2003). The Thing and I: Understanding the Relationship Between User and Product. In: M. Blythe & C. Overbeeke (Eds.) *Funology: From Usability to Enjoyment* (pp. 31–42). Dordrecht: Kluwer.
- Lindgren, A., Angelelli, A., Mendoza, P.A. & Chen, F. (2009). Driver behaviour when using an integrated advisory warning display for advanced driver assistance systems. *Intelligent Transportation Systems* 3(4), 390–399.
- Long, J. (1989). *Cognitive Ergonomics and Human Computer Interaction*. Cambridge University Press, Cambridge, GB.
- Mauney, D. W., & Masterton, C. (2008). Small-Screen Interfaces. In P. Kortum (Ed.), *HCI Beyond the GUI* (pp. 307–357). Burlington, MA: Elsevier.
- Moshagen, M., & Thielsch, M. T. (2011). Facets of visual aesthetics. *International Journal of Human Computer Studies*, 68(10), 689–709.
- Oulasvirta, A., Tamminen, S., Roto, V., & Kuorelahti, J. (2005). Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In G. C. van der Veer, C. Gale (Eds.), *Proceedings of CHI* (pp. 919–928). New York, NY: ACM.
- Seppelt, B.D. & Lee, J.D. (2007). Making adaptive cruise control (ACC) limits visible. *International Journal of Human-Computer Studies*, (65)3, 192–205.
- Sternberg, S. (1969). The discovery of processing stages: Extension of donder's method. *Acta Psychologica: Attention and Performance II*, 276–315.
- Stoner, H.A., Wiese, E.E., & Lee, J.D. (2003). Applying ecological interface design to the driving domain: The results of an abstraction hierarchy analysis. *Proceedings of the Human Factors Ergonomics Society*, 47, 444–448.
- Vicente, K. J., & Rasmussen, J. (1992). Ecological Interface Design: Theoretical Foundations. *IEEE Transactions on Systems, Man and Cybernetics* 22(4), 589–606.
- Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, S. (2012). *Engineering Psychology and Human Performance* (4th ed.). Boston: Pearson.
- Young, M.S., & Birrel, S.A. (2012). Ecological IVIS design: using EID to develop a novel in-vehicle information system. *Theoretical Issues in Ergonomics Science*, 13(2), 225–239.

AR und Webpad – nur Spiele(reien)?

Sven Buyer¹ und Carsten Wittenberg¹

Schlüsselwörter: Augmented Reality, Computerspiele, Mobile Endgeräte, industrielle Wartung

Zusammenfassung

Es wird Möglichkeiten zur Unterstützung von Wartungstechnikern mittels Anwendungen auf mobilen Endgeräten untersucht. Auf Basis einer Nutzeranalyse mit Wartungstechnikern wurden verschiedene Anforderungen, wie die Art der Informationsbeschaffung oder eine schnelle Informationsauswertung vor Ort aufgenommen.

Neben der Umsetzung als klassisches GUI wurden weitergehenden Visualisierungstechniken aus dem Bereich der Computerspiele untersucht. Als sinnvolle Visualisierungsmöglichkeit ist die Nutzung von Augmented Reality erkannt worden. Erweitert man das Realbild mit Zusatzinformationen aus dem aktuellen Kontext, bieten sich eine Reihe von Möglichkeiten zur intuitiven Unterstützung der Wartungstechniker wie Warn- oder Gefahrenmeldung.

Basierend auf der Nutzeranalyse und Game Design Pattern wird eine beispielhafte Anwendung auf einem Apple iPad zur Kommunikation mit einer Siemens S7-SPS umgesetzt. Relevante Design Pattern werden dabei aus der Welt der Computerspiele übernommen und auf die Anwendbarkeit in produktiven AR Anwendungen untersucht.

Abstract

The opportunities to support maintenance technicians with applications on mobile devices are analyzed. Based on an user analysis of maintenance technicians (n = 25), different requirements such as the type of information gathering or a quick evaluation of information have been determined.

Additional to the implementation of a classical GUI, the possibilities of further visualization techniques from the field of video games have been tested. Techniques of Augmented Reality may be used as a sensible way. By extending the real image with additional information from current context, a number of possibilities for intuitive assistance like warning signals facilitate the work of maintenance personnel.

Based on the user analysis and game design patterns, an application which communicates with a Siemens S7-PLC is implemented on an Apple iPad. Relevant design pattern are taken from the field of video games and are tested for the applicability to productive AR applications.

¹ Hochschule Heilbronn, Fakultät für Mechanik und Elektronik, 74081 Heilbronn

Anforderungen an AR-Anwendungen in der Instandhaltung

Ein Servicetechniker im industriellen Instandhaltungsbereich ist selbstständiges und eigenverantwortliches Teammitglied einer Arbeitsgruppe. Das grundlegende Tätigkeitsfeld erstreckt sich von Systemchecks, Fehleridentifikation und Störungsanalyse bis hin zur Problembehebung. Dies erfordert im Allgemeinen den Umgang mit Wartungs- und Anlagendokumentationen, Prozessdaten sowie die Kommunikation mit Remotespezialisten. Eine durchgeführte, explorative Umfrage (n=25 Personen) gibt Anlass dazu, die Art der Informationsbeschaffung und eine schnelle Auswertung dieser Informationen vor Ort, als ein elementares Kriterium einer effektiven Instandsetzung anzusehen. Dies bestätigen Antworten der Probanden auf die Frage nach dem prinzipiellen vorgehen beim Erhalt eines Auftrags zur Wartung einer Anlage mit den häufig vorkommenden Punkten „Technische Unterlagen zur Rate ziehen“, „Soll-Ist-Vergleich durchführen“ sowie „Bekanntmachen mit dem System“. Als problematisch stellen sich häufig eine dezentrale Lagerung der benötigten Unterlagen heraus. Dies führt zu zeitintensivem Suchen und Beschaffen der Dokumente, bevor effektiv mit der Arbeit an der entsprechenden Anlage begonnen werden kann.

Aufgrund der oben erläuterten Punkte der Dokumentenbeschaffung, Informationssuche sowie des anschließenden Transfers dieser Informationen auf realen Gegebenheiten, liegt der Schwerpunkt dieser Untersuchung auf der Informationsdarstellung und –übermittlung in produktiven AR Anwendungen.

Game Design Pattern zur Analyse digitaler Spiele

Heuristiken und Game Design Pattern für Computerspiele werden immer umfangreicher und detaillierter. Die von Korhonen (2007) beschriebenen Heuristiken sind im Rahmen eines iterativen Designprozesses eines Spieles für mobile Geräte entstanden. In Bezug auf die Anforderungen der durch Augmented Reality unterstützten Wartung, sind aus Sicht der Informationsbereitstellung, die Forderungen nach Integration der Menüs in das Gameplay sowie Bereitstellung von ausreichenden Ausgangsinformationen als wichtigste Punkte anzusehen. Zusätzlich finden sich hierauf basierende Ansätze zur Untersuchung der Beziehung zwischen Computerspielen und Anwendersoftware (Kallergi, 2010).

Game Design Pattern wiederum sind Muster zur Beschreibung von Spielmechanismen (Koster, 2005). Ziel ist es, das Spiel in Komponenten zu zerlegen, aus welchen Muster gewonnen werden können. Diese Mechanismen werden durch ein einheitliches Schema erfasst und bewertet. Sammlungen von Game Design Pattern erlauben den Vergleich von Konzepten und deren Auswirkungen auf das Game Play (Björk, 2006). Das Pattern „Information“ befasst sich mit dem Informationsfluss innerhalb von Computerspielen und den Möglichkeiten, dem Benutzer Informationen zugänglich zu machen. Unterkategorien setzten sich unter anderem mit Tooltips, Dialoggestaltungen oder Information Passing auseinander. Letzteres befasst sich mit der intuitiven Weitergabe von Informationen eines Charakters oder

Objektes innerhalb der Spielewelt und ist ein wichtiges Kriterium einer immersiven virtuellen Welt.

Ein triviales Beispiel ist die Verwendung von Avataren, welche häufig Informationen durch Änderungen ihres Aussehens, Verhaltens oder Position in impliziter Form oder als sichtbare Entitäten liefern. Sieht man weiterhin Menüoptionen und Funktionsauswahl als Objekte der Spielewelt, stellt eine kontextabhängige Darstellung dieser Objekte ebenfalls eine Informationsübermittlung dar. Es ist durchaus sinnvoll, dass der Spieler in Abhängigkeit der Benutzeroberfläche weiß, in welchem Kontext er sich gerade befindet und welche Optionen weiterhin zur Verfügung stehen.

Einen wirklichen Transfer zwischen Erkenntnissen aus der Computerspielforschung und dem Augmented Reality Bereich ist aktuell noch wenig zu erkennen. So konzentriert sich die Forschung im Bereich Interaktionsdesign für AR-Anwendungen hauptsächlich auf die Einblendung von Annotationen. Die Arbeiten von White (2007) und Zhang (2005) untersuchen den möglichen Einsatz virtueller Hinweise in Form von Text, Diagrammen, oder Animationen sowie die räumliche Orientierung und Zuordnung virtueller Informationen zu realen Objekten. Henderson und Feiner (2011) befassen sich unter anderem mit der Einblendung dreidimensionaler Werkzeuge als zusätzliche Unterstützung zu einigen Arbeitsschritten. Problematisch sind hier großflächige Verdeckungen der realen Welt und die korrekte Erzeugung räumlicher Tiefe.

Anwendung in Augmented Reality

Ein möglicher Ansatz der Informationssteuerung in Augmented Reality Anwendungen zeigt Abbildung 1. Hier wird mit Hilfe eines Avatars und dessen Änderungen die Aufmerksamkeit des Benutzers in der augmentierten Welt gesteuert. Abbildung 1 (links) zeigt den normalen Betriebszustand einer S7-Steuerung und den zugehörigen Avatar. Die Reaktionsänderung des Avatars zeigt dem Benutzer einen ungültigen Systemzustand an und lenkt die Aufmerksamkeit auf das entsprechende Objekt (Abb 1, rechts). Weitere Informationen können, einer dem betreffenden Objekt zugeordnete Visualisierungsfläche, entnommen werden. Hier symbolisiert ein Digitalmultimeter einen falschen Messwert. Somit entfällt das Auslesen und Decodieren eines Fehlercodes mit Hilfe zusätzlicher Werkzeuge.



Abb. 1: Augmented Reality Anwendung und S7-SPS. Links: Avatar im normalen Betriebszustand einer S7-Steuerung. Rechts: Reaktion des Avatars auf Messfehler.

Anschließend hat der Benutzer die Möglichkeit, das virtuelle Multimeter für weitere Informationen, auf die als fehlerhaft markierte Baugruppe zu ziehen (vgl. Abb. 2, links). Die Funktionsweise basiert auf dem aus Aufbau- und Simulationspielen bekannten Objekt-Objekt-Schema. Da hier aus der Vogelperspektive gespielt, in der Regel viele Objekte auf der Spielfläche vorhanden sind und diese entsprechend aufeinander angewendet werden können. Ein aktives Objekt (hier Multimeter) wird auf das passive Objekt „S7-Baugruppe“ angewendet. Durch Auswahl entsprechender Menüpunkte des aktiven Objektes können weitere Funktionen wie die grafische Darstellung eines Messwertes (vgl. Bild 2 rechts) ausgelöst werden.

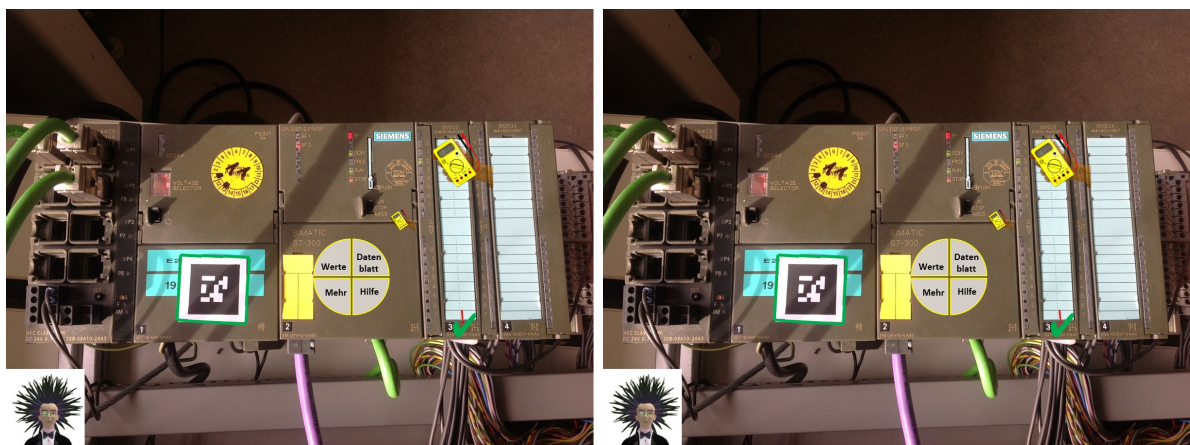


Abb. 2: Augmented Reality Anwendung und S7-SPS. Links: Menü zur Drag&Drop Aktion des Multimeters. Rechts: Visualisierung der Messwerte der aktiven Baugruppe.

Literatur

- Björk, S. & Holopainen, J. (2006) *Games and design patterns*. The Game Design Reader, MIT Press.
- Henderson, S. & Feiner, S. (2011). Exploring the Benefits of Augmented Reality Documentation for Maintenance and Repair. *IEEE Transactions on Visualization and Computer Graphics*, S. 1355 –1368.

- Kallergi, A. & Verbeek, F.J. (2010). Video games for collection exploration: games for and out of data repositories. *Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments*, York, NY, USA, S. 143–146.
- Korhonen, H. & Koivisto, (2007). Playability heuristics for mobile multi-player games. *Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts*, S. 28–35, Perth, Australia, ACM
- Koster, R. (2005). *A theory of fun for game design*. Paraglyph Press
- White, S., Lister, L. and Feiner, S. (2007). Visual Hints for Tangible Gestures in Augmented Reality. *6th International Symposium on Mixed and Augmented Reality ISMAR*, S. 47 –50, Nara, Japan
- Zhang, F. & Sun, H. (2005). Dynamic labeling management in virtual and augmented environments. *Ninth International Conference on Computer Aided Design and Computer Graphics*, S. 6, Hong Kong, China

Assistenz für Elektromobilität - Erfassung der Nutzerakzeptanz

Nicola Fricke und Katharina Oeltze¹

Schlüsselwörter: Elektromobilität, Nutzerakzeptanz, Usability, Fragebogen, Assistenz

Zusammenfassung

Im Rahmen dieses Beitrags wird ein Vorschlag für eine Item-Sammlung vorgestellt, der speziell die Akzeptanz von technischen Systemen zur Unterstützung der Nutzung von Elektromobilen erfasst. Hierfür wird sowohl auf bestehende Vorarbeiten im Bereich der Evaluation von Fahrerassistenzsystemen, als auch auf Arbeiten zum Technologie-Akzeptanz-Modell und bewährten „Usability“-Skalen zurückgegriffen. Die Item-Liste soll kontinuierlich weiterentwickelt, angepasst und im Rahmen der nationalen und internationalen Demonstrations- und Pilotprojekte zur Bewertung von Assistenzen und Systemen im Bereich Elektromobilität angewendet und erweitert werden.

Einleitung

Elektromobilität ist besonders im urbanen Raum ein aktuelles Thema. Vor allem im Rahmen von multi- und intermodalen Verkehrskonzepten wird die Nutzung von Elektrofahrzeugen als ein alternativer Verkehrsträger in der Mobilität der Zukunft eine maßgebliche Rolle spielen. Aktuelle nationale (z.B. Förderinitiative Schaufenster Elektromobilität) und internationale (z.B. EU CIP Projekt smartCEM, Mini-E Versuch von BMW) Demonstrations- und Pilotprojekte erkunden vor diesem Hintergrund Maßnahmen, mit denen eine Erhöhung der Verbreitung und Akzeptanz von Elektrofahrzeugen erreicht werden kann.

Bisherige Resultate der Nutzerbefragungen und Experteninterviews von diversen Projekte weisen darauf hin, dass die subjektive Wahrnehmung der Reichweite eine entscheidende Rolle bei der Nutzerakzeptanz spielt (Dimitropoulos, Rietveld, & van Ommeren, 2011; Götz, Sunderer, Birzle-Harder, & Deffner, 2011). Obwohl Studien zum allgemeinen Mobilitätsverhalten gezeigt haben, dass Nutzer selten weitere Strecken als 50km täglich fahren (Follmer et al., 2008), wird allgemein die Reichweite von Elektrofahrzeugen als unzulänglich eingestuft (z.B. Gärling, 2001). Ebenso zeigte eine Befragung im Rahmen des Förderschwerpunkts "Elektromobilität in Modellregionen", dass sich Nutzer vordringlich eine Verbesserung der Reichweite wünschen (Dütschke et al., 2012). Dieses Phänomen der Diskrepanz zwischen subjektiver Wahrnehmung und Einschätzung der benötigten Reichweite bzw. die Angst aufgrund mangelnder Reichweite das Ziel nicht zu erreichen, wird weitläufig als Reichweiten-Angst bezeichnet (siehe bspw. Gartner &

¹ Deutsches Zentrum für Luft- und Raumfahrt, Institut für Verkehrssystemtechnik

Wheelock, 2010). Ausdruck dieser Angst stellt die Nutzung von nur durchschnittlich 75 bis 80 % der eigentlichen Reichweite dar: außerhalb dieses Bereiches fühlen sich Nutzer zunehmend unwohl (Franke & Krems, 2013; Franke, Neumann, Bühler, Cocron, & Krems, 2012). Die Folge dieser Angst können Anpassungen im Mobilitätsverhalten sein. Die Mini-E Studie in den USA zeigte, dass die Nutzer ihr Mobilitätsverhalten anpassten und auf bestimmte Fahrten mit dem Elektrofahrzeug verzichteten (Turrentine, Garas, Lentz, & Woodjack, 2011). Inwiefern solche Anpassungen von den Nutzern toleriert werden, hängt sicher entscheidend davon ab, wie das "normale" Mobilitätsverhalten mit konventionellen Fahrzeugen aussieht und welche alternativen Transportmodi für den Nutzer verfügbar sind. So sinkt bspw. die Akzeptanz von voll-elektrischen Fahrzeugen, je schlechter die Zugänglichkeit zu öffentlichen Verkehrsmitteln eingeschätzt wird (Götz et al., 2011). Die Passung von Mobilitätsverhalten zu Kapazitäten des Elektrofahrzeugs scheint demnach eine relevante Stellgröße der Akzeptanz für geeignete, d.h. der Elektromobilität zugewandten, Zielgruppen zu sein (z.B. Peters & Hoffmann, 2011).

Die Entwicklung intermodaler und multimodaler Verkehrskonzepte ist hierbei ein guter Ansatz, um potentielle Nutzer mit einem bestimmten Mobilitätsmuster anzusprechen. Aktuelle Projekte, wie z.B. smartCEM, adressieren multimodale Reisekonzepte, in denen das Elektrofahrzeug ein Teil der Transportkette darstellt. Im Zuge dessen werden Informations- und Kommunikationstechnologien (IKT) und darauf aufbauende Assistenzen entwickelt, um die multimodale Einbindung verschiedener Transportmittel für den Nutzer einfach zu realisieren. Daneben sollen die entwickelten Applikationen die erwähnte Reichweitenangst adressieren, in dem sie Informationen über den Batteriestatus oder nächstmögliche Ladestationen geben. Prinzipiell sind IKT-Dienste und Anwendungen ein geeigneter Ansatzpunkt, um die Nutzung und die Akzeptanz von Elektrofahrzeugen zu erhöhen (siehe z.B. auch Peters & Popp, 2011).

Nur wenn derartige Systeme wiederum ansprechend, intuitiv und nutzerfreundlich gestaltet sind, lassen sich positive Effekte in der Gesamtakzeptanz von Elektrofahrzeugen erzielen. Dieser Bereich der Technologie-Akzeptanz wurde hierbei bereits in einer Reihe von Projekten beleuchtet. So wurde zum Beispiel das Technologieakzeptanzmodell von Davis (siehe Abbildung 1) zur Evaluation und Beurteilung technischer Assistenzsysteme schon mehrfach verwendet. Beispielsweise wurden im Rahmen des EU-Projekts eCoMove entsprechende Items zur Erfassung der „wahrgenommenen Nützlichkeit“ und zur „wahrgenommenen Bedienungs-freundlichkeit“ abgeleitet (Hörtl et al., 2012). Zur Erfassung der affektiven Komponente der „Einstellung zur Nutzung“ wurde die weitläufig Anwendung findende Skala zur Erfassung der Akzeptanz von Fahrzeug-Telematik benutzt (Van Der Laan, Heino, & De Waard, 1997). Speziell aus den Bereichen der „Usability“-Forschung liegen bewährte Erhebungsinstrumente vor (z.B. Brooke, 1986), die sich in angepasster Form auch für die Erhebung der wahrgenommenen Nützlichkeit

keit und Bedienungsfreundlichkeit im Kontext von Fahrzeugsystemen und somit auch elektromobilitätsspezifischen Diensten und Anwendungen eignen würden.

Zugrundeliegende Konzepte

Das Technologie-Akzeptanz-Modell nach Davis (1993) wurde bereits in einer Vielzahl von Studien zur Messung der Nutzerakzeptanz von technischen Systemen verwendet (siehe z.B. Bertrand & Bouchard, 2008; Cowen, 2009; Bargozi, & Warshaw, 1989; Gao, 2005).

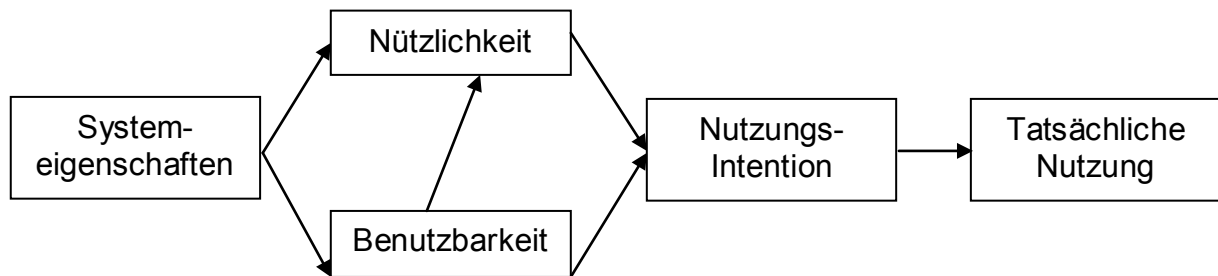


Abb.1: Technologie-Akzeptanz Modell (TAM - angepasst nach Davis, 1993).

Grundlage dieses Modells stellt die Theory of Reasoned Action (TRA) von Ajzen und Fishbein (1980) dar, die die kausalen Beziehungen zwischen Meinungen, Einstellungen, Verhaltensintention und tatsächlichem Verhalten beschreibt. So ist angelehnt an dieses Modell die Verhaltens- bzw. Nutzungsintention im Technologie-Akzeptanz-Modell der signifikanteste Prädiktor für die tatsächliche Nutzung eines Systems. Die Nutzungsintention wird dabei von zwei kognitiven Faktoren bestimmt: der wahrgenommenen Nützlichkeit und Benutzbarkeit eines Systems, was wiederum von Eigenschaften des Systems abhängt. Die wahrgenommene Nützlichkeit eines Systems stellt den Grad dar, zudem ein Individuum glaubt, dass die Nutzung eines Systems bei der Erledigung einer Aufgabe im Sinne einer Leistungssteigerung hilft. Die wahrgenommene Benutzbarkeit stellt dagegen den Grad dar, zudem ein Individuum glaubt, dass die Nutzung eines Systems frei von physischer und mentaler Anstrengung im Sinne einer einfachen Bedienbarkeit eines Systems ist. Abbildung 1 stellt das Modell dar.

Zur Operationalisierung der beiden Faktoren "Nützlichkeit" und "Benutzbarkeit" wurden bewährte Items des TAM (Davis, 1989), Items aus Usability-Skalen (z.B. Brooke, 1986; Lund, 2001; Moore & Benbasat, 2001) und weiteren Fragebögen aus Forschungsstudien (z.B. Hörtl et al., 2012) angepasst, ergänzt und zusammengefasst. Ziel ist es, diese Item-Sammlung als Basis zu verwenden und jeweils projektspezifische Anpassungen, z.B. bezüglich der Funktionalität der verwendeten Systeme vorzunehmen. Der Fragebogen stellt somit ein "work-in-progress" dar. Beispiele für die derzeit operationalisierten Skalen finden sich in Tabelle 1 und 2. Die Items sollen im Rahmen von Projekten zur Gestaltung und Bewertung von Assistenzsystemen im Anwendungskontext Elektromobilität verwendet und weiterentwickelt werden.

Tab.1: Items der Kategorie "Perceived-usefulness / Wahrgenommene Nützlichkeit"

Das System ist nützlich.
Das System hilft mir in der Erreichung meiner Fahrziele.
Das System gibt mir mehr Kontrolle bei der Nutzung des Fahrzeugs.
Das System hilft mir Zeit zu sparen.
Das System verhält sich so wie ich es erwarte.
Dank des Systems ist es leichter für mich, das Fahrzeug zu benutzen.
Das System hilft mir umweltfreundlicher zu fahren.
Das System ermöglicht stressfreieres Fahren.
Das System grenzt meine Freiheit beim Fahren ein.
Das System hilft mir, aktiv die Umwelt zu schonen.
Das System erhöht meine Bereitschaft das Fahrzeug zu nutzen.

Tab.2: Items der Kategorie "Perceived-ease-of use / Benutzbarkeit"

Die Interaktion mit dem System erfordert hohe mentale Anstrengung.
Meine Interaktion mit dem System ist klar und verständlich.
Das System ist rigide und unflexibel in der Nutzung.
Es ist für mich leicht, der Information des Systems zu folgen.
Ich finde das System ist umständlich zu nutzen.
Das System ist an meine individuellen Wünsche anpassbar.
Ich kann das System leicht dazu bringen das zu tun was ich möchte.
Ich finde das System ist leicht zu nutzen.
Ich finde das System ist nutzerfreundlich.
Ich finde das System unnötig komplex.

Literatur

- Ajzen, I., & Fishbein, M. (1980). *Understanding Attitudes and Predicting Social Behaviour*. Prentice-Hall, Englewood Cliffs, NJ.
- Bertrand, M., & Bouchard, S. (2008). Applying the technology acceptance model to VR with people who are favourable to its use. *Journal of Cyber Therapy and Rehabilitation*, 1(2).
- Bick, M., Kummer, T.-F., & Rössig, W. (2008). Ambient intelligence in medical environments and devices. Qualitative Studie zu Nutzenpotentialen ambienter Technologien in Krankenhäusern. Berlin.

- Brooke, J. (1986). SUS - A quick and dirty usability scale. Earley.
- Cowen, J. B. (2009). The Influence of Perceived Usefulness, Perceived Ease of Use, and Subjective Norm on the Use of Computed Radiography Systems: A Pilot Study.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, September, 318–340.
- Davis, F. D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies*, 38, 475–487.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management Science*, 35, 982–1003.
- Dimitropoulos, A., Rietveld, P., & van Ommeren, J. N. (2011). Consumer Valuation of Driving Range: A Meta-Analysis. *Tinbergen Institute Discussion Paper*, 133(3).
- Dütschke, E., Schneider, U., Sauer, A., Wietschel, M., Hoffmann, J., & Domke, S. (2012). *Roadmap zur Kundenakzeptanz: Zentrale Ergebnisse der sozialwissenschaftlichen Begleitforschung in den Modellregionen*. Karlsruhe: Fraunhofer ISI.
- Follmer, R., Gruschwitz, D., Jesske, B., Quandt, S., Lenz, B., Nobis, C., Mehlin, M. (2008). *MiD 2008 Mobilität in Deutschland 2008* (p. 214). Retrieved from http://www.mobilitaet-in-deutschland.de/pdf/MiD2008_Abschlussbericht_I.pdf
- Franke, T., & Krems, J. F. (2013). Interacting with limited mobility resources: Psychological range levels in electric vehicle use. *Transportation Research Part A: Policy and Practice*, 48, 109–122.
- Franke, T., Neumann, I., Bühler, F., Cocron, P., & Krems, J. F. (2012). Experiencing Range in an Electric Vehicle: Understanding Psychological Barriers. *Applied Psychology*, 61(3), 368–391.
- Gao, Y. (2005). Applying the Technology Acceptance Model to Educational Hypermedia: A Field Study. *Journal of Educational Multimedia and Hypermedia*, 14(3), 237–247.
- Gärling, A. (2001). Paving the Way for the Electric Vehicle.
- Gartner, J., & Wheelock, C. (2010). *Electric Vehicles: 10 Predictions for 2011*. Boulder: Pike Research LLC.
- Götz, K., Sunderer, G., Birzle-Harder, B., & Deffner, J. (2011). Attraktivität und Akzeptanz von Elektroautos: Arbeitspaket 1 des Projekts OPTUM: Optimierung der Umweltentlastungspotenziale von Elektrofahrzeugen. Frankfurt am Main: ISOE – Institut für sozial-ökologische Forschung.
- Hörtl, A., Trommer, S., Schießl, C., Fricke, N., Staubach, M., Hoek, R., Bernatallada, G. (2012). *Requirements and motivators for private and*

- commercial drivers*. D6.1. Retrieved from <http://ecomove-project.eu/assets/Documents/Deliverables/New/To-be-updated/120622-DEL-SP6-D6.1-updated.pdf>
- Lund, B. A. M. (2001). Measuring Usability with the USE Questionnaire. *STC Usability SIG Newsletter*, 8(2). Retrieved from http://www.stcsig.org/usability/newsletter/0110_measuring_with_use.html
- Moore, G. C., & Benbasat, I. (2001). Development of an instrument to measure the perceptions of adopting and information technology innovation. *Information Systems Research*, 2(3), 192–222.
- Peters, A., & Hoffmann, J. (2011). Nutzerakzeptanz von Elektromobilität Eine empirische Studie zu attraktiven Nutzungsvarianten, Fahrzeug- konzepten und Geschäftsmodellen aus Sicht potenzieller Nutzer. Karlsruhe: Fraunhofer ISI.
- Peters, A., & Popp, M. (2011). Elektroautos in der Wahrnehmung der Konsumenten Zusammenfassung der Ergebnisse einer Befragung in Deutschland. Karlsruhe: Fraunhofer ISI.
- Turrentine, T., Garas, D., Lentz, A., & Woodjack, J. (2011). *The UC Davis MINI E Consumer Study* (Vol. 0, pp. 1–78).
- Van Der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C*: 5(1), 1–10. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0968090X96000253>

The influence of user characteristics on the quality of judgment prediction models for tablet applications

Xin Guang Gong¹ and Klaus-Peter Engelbrecht²

Keywords: user characteristics, judgment predictability, tablet applications

Abstract

In this paper, we analyze the influence of five different user characteristics on the quality of judgment prediction models (multiple linear regression) for three Android tablet applications. The interaction parameters which we use as predictors in the models are based on the definitions in the ITU-T Recommendation Supplement 25 to P-Series (Möller, 2011). The following user characteristics have the strongest effects on the predictions: gender, age, technical affinity and technical experience.

Introduction

User interfaces (UIs) nowadays are designed for user experience, which implies to achieve high hedonic and pragmatic quality (Hassenzahl, Burmester & Koller, 2003). While hedonic qualities are measured with questionnaires only, pragmatic quality is often measured with a mix of questionnaires and instrumental measures. Since user judgments and instrumental measurements can be contradictory (Frøkjær, Hertzum & Hornbæk, 2000), it is interesting to study their relation.

One approach to study this relation is statistical modeling, e.g. utilizing multiple linear regression (MLR, figure 1), in which instrumental measures of an interaction are used as predictors, and the user judgment of the same interaction is used as a dependent variable (Walker, Litman, Kamm, & Abella, 1997). Such a model can also be useful to predict user judgments when the user cannot be asked, e.g. at certain events during the interaction, or when instrumental data are collected in the field.

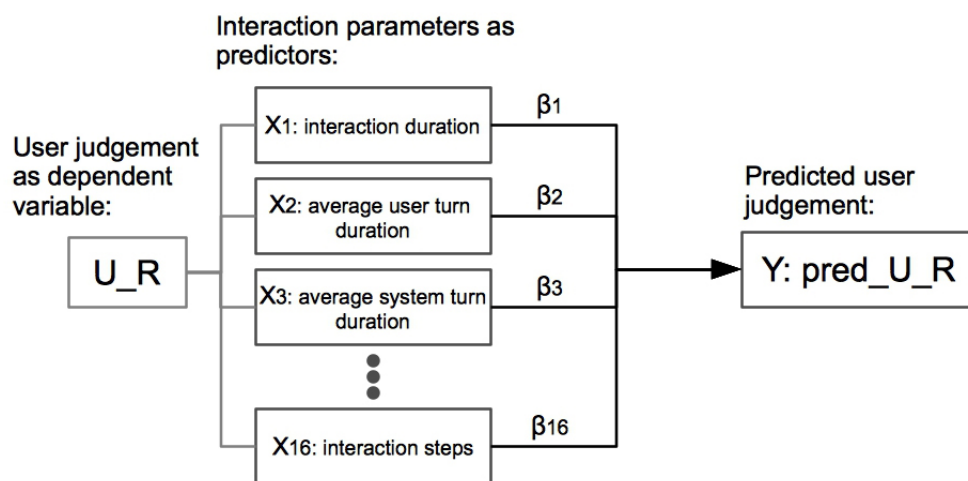
Since different types of users differ in their interaction behavior and in their requirements towards user interfaces (e.g. Nielsen, 1993), it is reasonable to also assume differences in the relation between judgments and instrumental measurements for different types of users. This was analyzed in a previous study (Engelbrecht, Möller, Schleicher & Wechsung, 2008) with a speech-based smart-home system. The study included 16 younger (20-30y) and 16 older (60-80y) adults, of which technical affinity and memory performance was measured. Fourteen tasks with instrumental measures and quality judgments were available for

¹ Quality and Usability Lab, Telekom Innovation Laboratories, Technische Universität Berlin
Ernst-Reuter-Platz 7, 10587 Berlin, xin-guang.gong@telekom.de

² Quality and Usability Lab, Telekom Innovation Laboratories, Technische Universität Berlin
Ernst-Reuter-Platz 7, 10587 Berlin, klaus-peter.engelbrecht@telekom.de

each user. It was shown that users mainly differed in how many and how strongly instrumental measures were correlated with the judgments, but hardly differed in the type of correlated parameters. Moreover, it was shown that MLR models for the 50% users with best memory performance explained more of the variance in the user judgments than a model for the worst 50%. It was concluded, that users may have tried to judge the interactions in accordance with instrumental measures, but differed in their ability to do so.

This paper presents results from a new study which was performed to confirm the results of Engelbrecht et al. (2008), but this time using touch-based graphical user interfaces (GUIs) instead of a voice user interface. Since no prior study on predicting user judgments from instrumental measures for GUIs is known to us, a further aim of the study was to analyze the applicability of the approach to such systems.



Multiple Linear Regression(MLR):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_{16} X_{16i} + \epsilon_i$$

$$i = 31 \text{ users} * 11 \text{ tasks} = 341$$

Fig.1: MLR model for one task

Experiment and Users

The design of the experiment was similar to the design used by Engelbrecht et al. (2008). 16 Younger users (17-27y, $M=21.9$, $SD=3.7$; 7 female, 9 male) and 15 older users (59-84y, $M=70.1$, $SD=7.5$; 9 f, 6 m) were recruited at the TU Berlin campus and in supermarkets nearby. Most of the younger participants were students of TU Berlin; the older participants had different professions. All participants had good to very good German language comprehension abilities. Over 96% (30 of 31) of all participants had a mobile-phone and over 58% (18 of 31) had a smart-phone. We expected a clear difference of the scores in the memory test in both age groups.

Each user performed 3-4 tasks with each of three GUI applications running on a tablet with touch screen (11 tasks overall). Prior to the experiment, the tablet was explained to the users briefly. Also, to maximize chances that especially the older users would be able to perform all tasks without help, each app was explained to the users briefly, and tasks within one application were sorted so that users would gradually learn how the interface works. However, the order of apps was randomized.

At the beginning of the experiment, we collected information about the users' technical affinity, using the same scale as used by Engelbrecht et al. 2008, their experience with touchscreen devices and their gender and age. At the end of the experiment, memory performance was tested using the digit span test of the Wechsler intelligence test (German version; von Aster, Neubauer & Horn, 2006).

After each task, the participants rated the suitability of the application for this task, using the scale displayed in Figure 2 (U_R in the following). In addition, cognitive effort was self-assessed using the SEA scale (Skala zur Erfassung subjektiv erlebter Anstrengung; Eilers, Nachreiner & Hänecke, 1986). At the end of each application, the participants filled out the AttrakDiff mini questionnaire (Diefenbach & Hassenzahl, 2010).

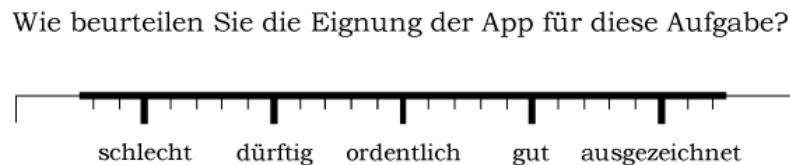


Fig.2: U_R rating scale

All interactions were logged using the *Android HCI Extractor* tool (Mateo Navarro & Hillmann, 2012). The concepts of this tool are based on prior work of Möller (2011). Within the logs, the interaction is divided into system turns and user turns. Each turn includes a delay, feedback and action stage. At each stage, time stamps and meta-data about the interaction process (e.g. number of system UI elements, or number of elements provided by the user) are collected.

This allowed us to calculate different types of interaction parameters for each task. These cover several aspects of the interaction: *interaction efficiency* (e.g. dialog duration), *cognitive efforts* (e.g. user feedback delay), *executing efforts*, *application complexity* (e.g. number of system UI elements) and *input efficiency* (e.g. action stage duration). Table 1 shows all 16 interaction parameters.

Tab.1: Interaction parameters

interaction duration	average user turn duration	average system turn duration	average user feedback delay
average system feedback delay	maximum system feedback delay	maximum user feedback delay	average user action duration
average system action duration	average system concepts	maximum user elements	average user text elements
average system elements	average user elements	sum of user transferring actions	interaction steps

Modeling

In order to examine the influences of a user characteristic on the performance of judgment prediction models, we divide the user sample into two groups (at the median in case of digit-span (Median=19), technical affinity (Median=3.86) and experience (Median=6)). Then, we calculate models to predict the U_R judgments for each subgroup, and compare the explained variance (R^2) using t -test. In addition, the root-mean-squared error ($RMSE$) is calculated for each subgroup.

We use multiple linear regression as a modeling algorithm. In order to get an estimate of the prediction performance of the models, the R^2 and $RMSE$ of each subgroup are calculated by 10-fold cross-validation (CV). Ten repetitions of CV are performed to balance out potential biases due to sampling the CV partitions. The distributions of the R^2 -values for the ten repetitions are then compared using t -test. The selection of the instrumental parameters used as predictors follows a machine learning approach: In each iteration of the CV, the predictor set performing best on the training data is determined. Then, a model is trained with all training data and this predictor set. This model is used to predict the cases in the current test partition.

This procedure is repeated for all five user characteristics measured in our study: gender, age, memory abilities, technical experience and technical affinity. In comparison, Engelbrecht et al. (2008) had only measured age, memory abilities and technical affinity.

Before we analyze the influence of user characteristics on the judgment prediction models, we analyze the correlation between the user characteristics. Table 2 shows a correlation matrix of the four user characteristics for all 31 participants. The age has a strong correlation to technical experience and affinity, but surprisingly it does not have a strong correlation to the digit-span score. Technical experience and technical affinity are strongly correlated with each other.

Tab.2: Correlation matrix for age, digit-span, technical experience and affinity

N=31	Age (r value / p value)	Digit-span (r value / p value)	Techn. Exp. (r value / p value)	Techn. Affi. (r value / p value)
Age	1	-0.26 / 0.16	-0.87 / 0.00	-0.58 / 0.00
Digit-span	-0.26 / 0.16	1	0.32 / 0.08	0.39 / 0.03
Techn. Exp.	-0.87 / 0.00	0.32 / 0.08	1	0.60 / 0.00
Techn. Affi.	-0.58 / 0.00	0.39 / 0.03	0.60 / 0.00	1

Results

In Table 3, we report the mean R^2 and $RMSE$ over the ten CV repetitions of each subgroup. The $RMSE$ shows that the prediction error amounts to roughly one judgment level on the U_R rating scale (one level is equal to 5 scale units). The influences of each user characteristic (gender, age, digit-span, technical experience and technical affinity) on the performance of the judgment prediction model can be seen by the difference in the R^2 values for the respective sub-groups. For women, older participants and participants with lower technical experience or lower technical affinity, the prediction accuracy is significantly better than for the respective complementary groups. Only for the digit-span subgroups, the R^2 values are very similar and do not differ significantly ($U_R=0.14$ compared to 0.13).

**Tab.3: R^2 and $RMSE$ calculated for different sub-groups of the database,
*($p<0.05$ for t -test comparing mean R^2 in ten CV repetitions for both subgroups)**

sub-groups	R^2	$RMSE$
Gender: male/female	0.07 / 0.23*	4.14 / 4.46
Age: young/old	0.03 / 0.18*	3.83 / 5.09
Digit-span: low / high	0.14 / 0.13	4.73 / 3.89
Tech. Affi.: low / high	0.28* / 0.04	4.46 / 4.26
Tech. Exp.: low / high	0.23* / 0.00	5.09 / 3.93

In order to analyze how similarly the ratings of individual users are influenced by the aspects of the interaction which we measured instrumentally, we calculate user-wise correlations between the U_R ratings and interaction parameters for the 11 tasks performed by each user. We then compare the significant ($p<0.05$) correlations found for each user.

As in the study of Engelbrecht et al. (2008), the number of significant correlations differs between the users. Over 80% (25 of 31) of our participants are showing correlations between their judgments (U_R) and interaction parameters. For some users, up to 8 parameters are correlated with the judgments. Although generally the parameters correlated with the user judgments are more diverse than in the study of Engelbrecht et al. (2008), there are some which appeared more frequently: *average system turn duration* (10 of 31 users), *average system action duration* (10

of 31 users), *average system elements* (6 of 31 users), *maximum user feedback delay* (6 of 31 users) and *interaction duration* (6 of 31 users). The first three of these parameters are related to the application's complexity, the other two are related to cognitive effort and time efficiency, respectively. *Average system turn duration* and *interaction duration* were also often correlated with user judgments in the study of Engelbrecht et al. (2008).

Discussion and Conclusion

In this paper, we described a study aimed at confirming results of an earlier study in which it was shown that user characteristics, and especially memory performance, are crucial for the performance of models for the prediction of user judgments from instrumental measures of the interaction. However, while in the previous study a voice user interface was used, this study analyzed interactions with touch-based tablet applications.

We could show that the user characteristics gender, age, technical affinity and technical experience all make a difference for the prediction performance. However, unlike in the prior study, memory performance had no influence. The results of the digit-span test are in both groups quite similar (young_avg=19.9, old_avg=18.3), so the two groups obtained from splitting at the median may have been too similar. In addition, the digit-span test is just a small part of the participant's cognitive abilities. Maybe in our generation it became unusual for young people to remember and repeat a long sequence of numbers. Or the digit-span test may not be a good indicator for the memory requirements in this experiment. While the digit-span test may be a reasonable measurement for the user's cognitive ability in a spoken dialog system, for a tablet application with a graphical user interface, another type of measurement may be more useful, e.g. a visual memory test.

The observed differences for the other user characteristics are obviously not caused by different abilities of the users to judge consistently throughout the experiment. Possibly, the less well predictable user groups focused on aspects of the interaction which were not represented in our instrumental measurements when judging the suitability of the app for the tasks. Since instrumental measures mostly address pragmatic aspects of the interaction, women, elderly, less technical affine and less technically experienced users may focus more on pragmatic aspects than the respective complementary subgroups. This may be due to differences in preferences, requirements or task difficulty between the subgroups. In follow-up evaluation of the data, this will be analyzed using the SEA and AttrakDiff ratings.

Generally, it could be shown that prediction of user judgments is somewhat possible using interaction parameters derived from the logs of the Android HCI extractor tool. Our next step in this respect is adding more interaction parameters, which fit more precisely our needs, e.g. the quantity of operating errors or we ex-

tent the logging framework for all interactions with the tablet device and not only for the interactions with the GUI. With those changes we might be able to increase our prediction accuracy and by this contribute to the automatic usability evaluation approaches.

Literature

- Von Aster, M., Neubauer, A., & Horn, R. (Ed.) (2006). WIE. Wechsler Intelligenztest für Erwachsene. Übersetzung und Adaption des WAS-III von David Wechsler, USA: The Psychological Corporation.
- Diefenbach, S. & Hassenzahl, M. (2010). *Handbuch zur Fun-ni Toolbox. User experience Evaluation auf drei Ebenen*. Retrieved from http://fun-ni.org/wp-content/uploads/Diefenbach+Hassenzahl_2010_HandbuchFun-niToolbox.pdf
- Eilers, K., Nachreiner, F. & Hänecke, K. (1986). Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Anstrengung. *Zeitschrift für Arbeitswissenschaft*, 40, 215-224.
- Engelbrecht, K.-P., Möller, S., Schleicher, R. & Wechsung, I. (2008). *Analysis of PARADISE Models for Individual Users of a Spoken Dialog System*. In Proceedings of ESSV 2008, Frankfurt/Main, Germany, 86-93.
- Frøkjær, E., Hertzum, M. & Hornbæk, K. (2000). *Measuring usability: are effectiveness, efficiency, and satisfaction really correlated?* In Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '00). New York, NY, USA: ACM, 345-352.
- Hassenzahl, M., Burmester, M. & Koller, F. (2003) AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In J. Ziegler & G. Szwillus. *Mensch & Computer 2003. Interaktion in Bewegung*, 187-196, Stuttgart, Leipzig: B.G. Teubner.
- Mateo Navarro, P.L. & Hillmann, S. (2012). Model-based Measurement of Human-Computer Interaction in Mobile Multimodal Environments. *AMMI 2012*, 30-33.
- Möller, S. (2011). Parameters Describing the Interaction with Multimodal Dialogue Systems. ITU-T Recommendation Supplement 25 to P-Series, Geneva, Switzerland: International Telecommunication Union.
- Nielsen, J. (1993). *Usability Engineering*, San Diego, CA, USA: Academic Press.
- Walker, M.A., Litman, D.J., Kamm, C.A. & Abella, A. (1997). PARADISE: a framework for evaluating spoken dialogue agents. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics (ACL '98), 271-280, Stroudsburg, PA, USA: Association for Computational Linguistics.

ACT-CV: Bridging the Gap between Cognitive Models and the Outer World

Marc Halbrügge¹

Keywords: Cognitive Modeling, Human Computer Interaction, Usability, Prototyping

Abstract

Cognitive modeling can provide important insights about the usability of software or hardware products at all stages of product design. It is not used very often, though, as the creation of cognitive user models is typically done by highly trained experts. These specialists need not only develop the models, but have to spend a lot of additional time on software interface problems: The cognitive models need to be connected to the interactive system under evaluation. This paper presents ACT-CV, a software library that aims at solving the interface problems of user modeling especially in usability research.

Introduction

Cognitive modeling has proven more and more useful for the design and evaluation of human machine interfaces as it allows to predict usability metrics like the time to complete a task or the types of errors that may occur. The psychological basis for cognitive user models is laid by cognitive architectures like ACT-R (Anderson, 2007). ACT-R is a software framework that provides a well-researched environment that closely mimics the cognitive, perceptual, and motor capabilities of human users. Unfortunately, cognitive architectures tend to lock the modeler into highly specialized computer systems (Lisp in the case of ACT-R), creating significant hurdles for the application of a user model to the human machine interface in question (Byrne, 2013).

In the case of basic psychological research, like visual search experiments, the user interface (UI) can be created within the system, e.g. implemented using Lisp (see also Urbas et al., 2006). This is not an option for existing software that already has been developed, and where Lisp is highly uncommon or in cases where changes to the original software are not possible.

CogTool (John et al., 2004) is an extremely powerful tool for creating user models, but it only works on mock-up prototypes of the user interface under evaluation, i.e. it only eases the lock-in problem, but doesn't solve it. A different approach is taken by SegMan (Ritter et al., 2006) which identifies elements on the screen based on pre-defined pixel color comparisons. SegMan's major advantage is that it al-

¹ Quality and Usability Lab, Telekom Innovation Laboratories, Technische Universität Berlin
Ernst-Reuter-Platz 7, 10587 Berlin, marc.halbruegge@telekom.de

lows to break out of the Lisp system, but its screen processing capability is rather hard to use and cannot handle dynamic screens very well.

The software library ACT-CV (Halbrügge et al., 2007) uses a similar approach as SegMan for allowing direct interaction between ACT-R and arbitrary software. It differs a lot on the technical side, as it uses computer vision for the identification of screen elements and adds support for time based visual parameters like optical flow. The graphic processing inside ACT-CV is handled by the open source framework OpenCV (<http://opencv.org/>) which provides state of the art machine vision on multiple platforms and includes more than 2500 individual algorithms.

The computer vision approach allows the extraction of much more complex features of the computer screen than just the color of a given pixel. In addition, it provides access to the external world using a video camera.

ACT-CV was used for modeling visual tasks especially in the automotive domain, where it allowed to connect a cognitive model of car driving (Salvucci, 2006) to an already existing driving simulation (Deml and Neumann, 2008; Halbrügge et al., 2008, see figure 1). The applications of ACT-CV, while successful, have also brought some weaknesses and other issues into light. This paper is about how these issues are addressed by an upcoming new revision of the software library.



Figure 1: Driving simulation experiment from Halbrügge et al. (2008)

Application and Shortcomings of ACT-CV 1

The simple task of finding the position of a predefined visual element (e.g. a search button) on the computer screen can be accomplished by using ACT-CV out of the box and providing nothing but a picture file containing that visual element to it.

But when the task becomes more complex, the usability of ACT-CV degrades sharply. A potential user of the system has to extend the software with his or her own machine vision code in order to apply it to new interfaces. While recompiling the system needs only average software development skills, designing the sensor code that provides information from the outer world to the cognitive model demands some knowledge of computer vision algorithms as well.

Another problem of ACT-CV 1 is the lack of text processing. Most interfaces rely at least partly on written language. ACT-CV in its first incarnation comes with many sensors for colors and shapes, but it cannot decipher text. This shortcoming is hard to overcome, e.g. the sensor for the visual search on road signs task in Halbrügge et al. (2008) had to be implemented as a comparison of pictures of letters, which is an extremely CPU intensive and error prone workaround. Figure 2 demonstrates the signal flow that was used in this experiment.

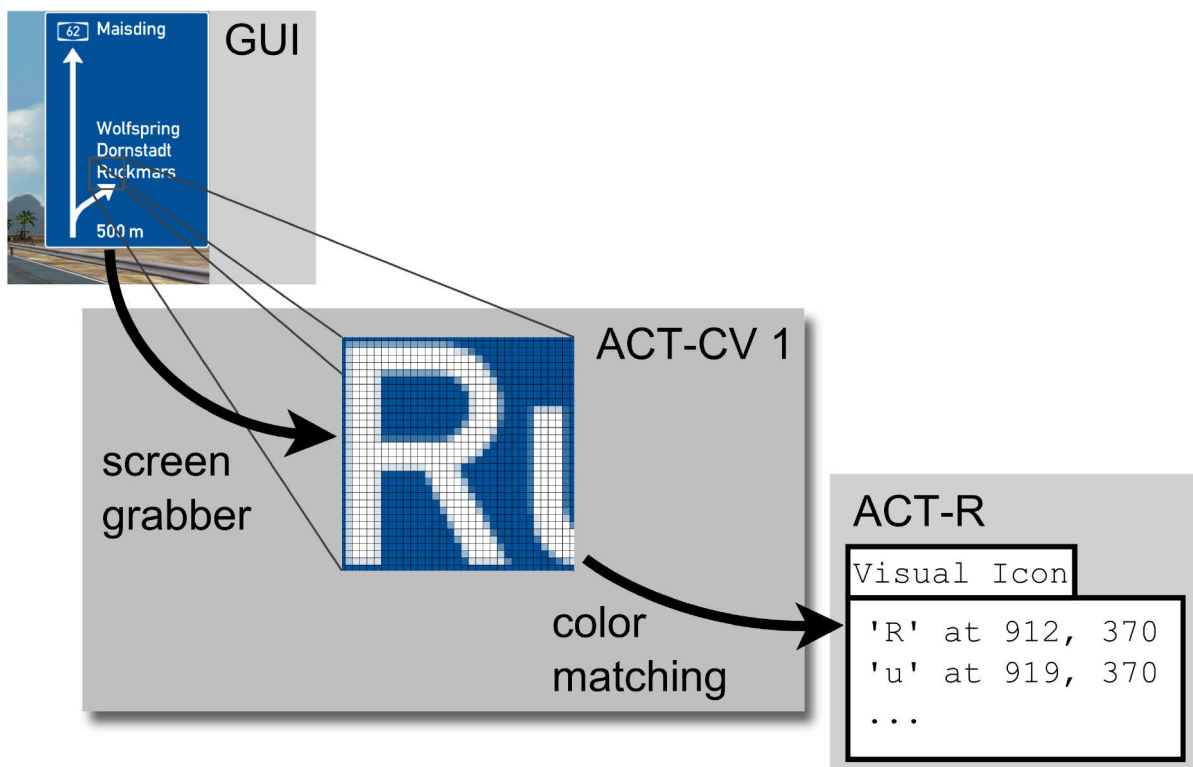


Figure 2: Application of ACT-CV 1 to the road sign reading task

ACT-CV 2

ACT-CV 2 is a major overhaul of the system. It simplifies the use of the library and at the same time addresses the lack of text processing in the previous revision of ACT-CV by adding HTML5 support. HTML5 (Hickson and Hyatt, 2011) marks the progression of HTML from a markup language for static text documents to a programming language for interactive user interfaces. It powers not only the web, but many native mobile applications, and is used as platform for rapid prototyping

as well. ACT-CV 2 features a state of the art web browser that allows both human subjects and cognitive models to interact with any HTML5 application.

It uses the open source WebKit (<http://www.webkit.org/>) rendering engine for both creating the GUI that is presented to the user and for extracting the (textual) features that are handed over to ACT-R (see Figure 3). The pixel based computer vision approach of ACT-CV 1 is still available, allowing the combination of symbolic text features extracted from the HTML document (DOM) with sub symbolic visual features like bottom up salience or optical flow extracted from the rendered user interface.

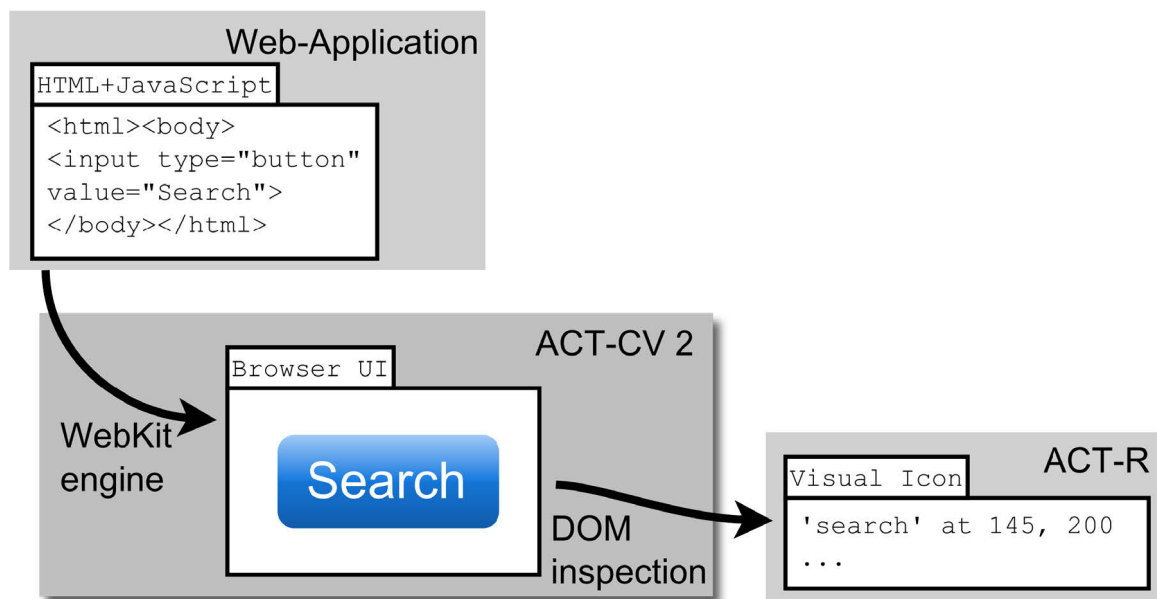


Figure 3: Signal flow for HTML5 applications in ACT-CV 2

Planned Applications of ACT-CV 2

The first application of ACT-CV 2 will be the usability evaluation of an HTML based cooking assistant that has been developed as part of the model based user interface project MASP (Blumendorf et al., 2008). ACT-CV will be used in conjunction and comparison with the MeMo workbench for automated usability evaluation (Engelbrecht et al., 2009; Quade et al., 2013).

MASP automatically adapts the UI of the cooking assistant to the different devices that it supports (e.g. smartphone, TV). This creates the need to be able to test the usability of the UI for each variation in layout and size. Covering all those variations using CogTool or SegMan would create significant overhead as one had to address all of them individually. ACT-CV 2 on the other hand can handle all variations at once as long as the wording on the screens doesn't change.

Future Extensions

Future extensions of ACT-CV will probably include automatically generated saliency maps based on the graphical input to the system. This will form the basis for

more accurate models of visual search, as needed for example to predict the probability to overlook a button in a given UI (Ruß, 2011). The hybrid approach of ACT-CV 2 is well suited for this extension as OpenCV already comes with bio-inspired algorithms for foveal vision (Benoit et al., 2010).

Another interesting extension is the integration of Java ACT-R (Salvucci, 2010). As ACT-CV uses Java for accessing the screen and translating mouse and keyboard events generated by the cognitive architecture to actual events of the operating system, the inclusion of Java ACT-R should not create too much effort. It would substantially ease the creation and maintenance of cognitive user models.

Conclusions

In order to use cognitive modeling during the usability engineering process, one has to combine the model on the one hand with the interactive environment of the software in question on the other hand. Creating interfaces between both is technically challenging. ACT-CV, the software library presented here, aims at bridging the gap between cognitive models and the outer world. It provides access to any graphical user interface by the means of machine vision applied to the current content on the computer screen. But the universality of ACT-CV's approach comes with the downside of demanding basic software engineering and computer vision knowledge.

ACT-CV 2, the upcoming revision of the library, adds new features while easing its use by adding native support for HTML5 applications. In the special case of HTML based user interfaces, ACT-CV 2 allows to bypass the computer vision layer and provides the textual elements of the UI directly to the vision module of the cognitive architecture. This new approach renders laborious tasks like manually annotating screenshots of the UI unnecessary and allows the modeler to fully concentrate on the cognitive user model itself.

ACT-CV is available for download at <http://act-cv.sourceforge.net>

References

- Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford University Press, New York.
- Benoit, A., Caplier, A., Durette, B., and Hérault, J. (2010). Using human visual system modeling for bio-inspired low level image processing. *Computer vision and Image understanding*, 114(7):758–773.
- Blumendorf, M., Feuerstack, S., and Albayrak, S. (2008). Multimodal user interfaces for smart environments: the multi-access service platform. In *Proceedings of the working conference on Advanced visual interfaces*, pages 478–479. ACM.

- Byrne, M. D. (2013). Computational cognitive modeling of interactive performance. In Lee, J. D. and Kirlik, A., editors, *The Oxford Handbook of Cognitive Engineering*, chapter 27. Oxford University Press, New York.
- Deml, B. and Neumann, H. (2008). Predicting the distraction of car drivers by cognitive modelling. *ATZe Elektronik*, 3(2):14–17.
- Engelbrecht, K.-P., Quade, M., and Möller, S. (2009). Analysis of a new simulation approach to dialog system evaluation. *Speech Communication*, 51(12):1234–1252.
- Halbrügge, M., Deml, B., Färber, B. A., and Bardins, S. (2007). ACT-CV – Die Erweiterung von ACT-R um Bildverarbeitungsalgorithmen erlaubt die schnelle Erzeugung mächtiger Benutzermodelle. In Grandt, M. and Bauch, A., editors, *Simulationsgestützte Systemgestaltung – DGLR-Bericht 2007-04*, pages 313–331, Bonn. DGLR e.V.
- Halbrügge, M., Deml, B., and Neumann, H. (2008). A cognitive model of visual search on direction signs alongside the autobahn. *International Journal of Psychology*, 43(3/4):759.
- Hickson, I. and Hyatt, D. (2011). *HTML5: A vocabulary and associated APIs for HTML and XHTML*. W3C Working Draft edition.
- John, B. E., Prevas, K., Salvucci, D. D., and Koedinger, K. (2004). Predictive human performance modeling made easy. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 455–462. ACM.
- Quade, M., Lehmann, G., Engelbrecht, K.-P., Roscher, D., and Albayrak, S. (2013). Automated usability evaluation of model-based adaptive user interfaces for users with special and specific needs by simulating user interaction. In Martin, E., Haya, P. A., and Carro, R. M., editors, *User Modeling and Adaptation for Daily Routines: Providing Assistance to People with Special Needs*, pages 219–247. Springer, Berlin / Heidelberg.
- Ritter, F. E., Van Rooy, D., Amant, R. S., and Simpson, K. (2006). Providing user models direct access to interfaces: an exploratory study of a simple interface with implications for hri and hci. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 36(3):592–601.
- Ruß, A. (2011). Modeling Visual Attention for Rule-based Usability Simulations of Elderly Citizen. In Harris, D., editor, *Engineering Psychology and Cognitive Ergonomics*, pages 71-81. Springer, Berlin.
- Salvucci, D. D. (2006). Modeling driver behavior in a cognitive architecture. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(2):362–380.
- Salvucci, D. D. (2010). *Cognitive supermodels*. Invited talk, European ACT-R Workshop, Groningen, The Netherlands.
- Urbas, L., Heinath, M., Trösterer, S., Pape, N., Dzaack, J., Kiefer, J., and Leuchter, S. (2006). Agimap: A tool-chain to support the modeling of the interaction

level of dynamic systems. In *Proceedings of the 7th International Conference on Cognitive Modeling* (S. 409). Trieste, Italien: Erlbaum.

Mental Models und Image Schemata: Gestaltung komplexer Software mit Hilfe von mentalen Modellen der Nutzer

Annika Johnsen¹, Diana Löffler² und Matthias Rötting¹

Schlüsselwörter: Mental Models, Image Schemata, User Centered Design, Softwaregestaltung

Zusammenfassung

Die Berücksichtigung der mentalen Modelle der Nutzer von der Interaktion mit einem Produkt trägt zur Gestaltung gebrauchstauglicher und intuitiv benutzbarer Software bei. Die Mental Models Methode (Young 2008) liefert in diesem Zusammenhang eine Gegenüberstellung der Nutzerperspektive mit den funktionalen Elementen einer Software. Dabei können sogenannte „Gaps“ aufgedeckt werden: Motivationen und Aufgaben der Nutzer, die nicht in angemessener Weise durch Funktionen einer Software unterstützt werden. Die Methode der Image Schemata (Hurtienne 2011) ergänzt die Mental Models Methode, indem zu jenen Gaps konkrete Gestaltungshinweise in Form von image-schematischen Metaphern (Hurtienne & Blessing 2007) generiert werden, sodass bei Kombination beider Methoden Gestaltungslücken nicht nur aufgedeckt, sondern auch geschlossen werden können. Im folgenden Beitrag wird die praktische Umsetzung beider Methoden am Beispiel der Bildbearbeitungssoftware GIMP erläutert. Unter anderem wurden einige der durch image-schematische Metaphern generierten Gestaltungshinweise umgesetzt, in einen Low-Fidelity Prototypen eingefügt und durch Nutzertests evaluiert.

Nutzerzentrierte Gestaltung komplexer Softwaresysteme

Softwaresysteme mit komplexem Funktionsumfang können einer technischen Entwicklung zugeschrieben werden, die es ermöglicht, immer mehr Funktionen und Optionen in die Gestaltung einer Software miteinzubeziehen, sodass der Nutzer durch zahlreiche Elemente in seinen Tätigkeiten unterstützt wird. Allerdings erfordert die Benutzung derartiger Softwaresysteme meist ein intensives Auseinandersetzen des Nutzers mit den verschiedenen Elementen der Nutzungsoberfläche. So stellt sich die Frage, nach welchen Kriterien Software zu gestalten ist, die einerseits durch einen großen Funktionsumfang den Anforderungen ihrer Nutzer entspricht, andererseits weiterhin "einfach" zu benutzen ist. Dabei sollte die Integration neuer technischer Innovationen in die Nutzungsoberfläche einer Software nicht im Widerspruch zum nutzerfreundlichen Einsatz und Nutzungsbedingungen stehen.

¹ Technische Universität Berlin, Fachgebiet Mensch-Maschine-Systeme

² Universität Würzburg, Lehrstuhl für Psychologische Ergonomie

Um die Nutzungsoberfläche komplexer Software für den Nutzer zugänglich und nachvollziehbar zu gestalten, eignet sich eine benutzerzentrierte Herangehensweise nach ISO 9241-210. Der Begriff bezieht sich auf "Methoden, welche auf die Gestaltung für und die Miteinbeziehung von Nutzern bei der Entwicklung computer-gesteuerter Systeme abzielen" (Abrás et al. 2004). Durch die Miteinbeziehung des Nutzers in den Gestaltungsprozess werden Effektivität, Effizienz und Sicherheit bei der Interaktion mit einem Produkt erhöht und die Kundenzufriedenheit gesteigert (Abrás et al. 2004). Ein Teilkonzept nutzerzentrierter Gestaltung stellt das Konzept der intuitiven Benutzbarkeit dar. Dabei steht der Begriff für das "Ausmaß, in dem die Benutzung eines Produktes auf der unbewussten Anwendung von Vorwissen beruht und dadurch zu einer effektiven und zufriedenstellenden Interaktion bei minimalem Verbrauch kognitiver Ressourcen führt" (Mohs et. al. 2006).

Methodische Grundlagen

Im vorliegenden Beitrag werden zwei Methoden nutzerzentrierter Gestaltung anhand eines praktischen Beispiels miteinander verknüpft. Dabei wurde zunächst die Mental Models Methode nach Young (Young 2008) auf das Bildbearbeitungsprogramm GIMP angewendet und anschließend durch die Methode der Image Schemata nach Hurtienne (Hurtienne 2011) ergänzt. Im Hinblick auf Software kann die Mental Models Methode beschrieben werden als visuelle Gegenüberstellung der Motivationen und Vorstellungen, die der Nutzer von einer Software hat (Mental Map) und der Funktionen, die diese dem Nutzer zur Verfügung stellt (Content Map). Die Nutzerperspektive wird durch non-direktive Interviews erfasst, bei denen die Nutzer sich zu verschiedenen, anwendungsbezogenen Aspekten der Software äußern. Das Sprachmaterial wird anschließend nach sogenannten Atomic Tasks durchkämmt: Bedürfnisse, Motivationen und Verhaltensweisen, die im Zusammenhang mit der Software geäußert werden.

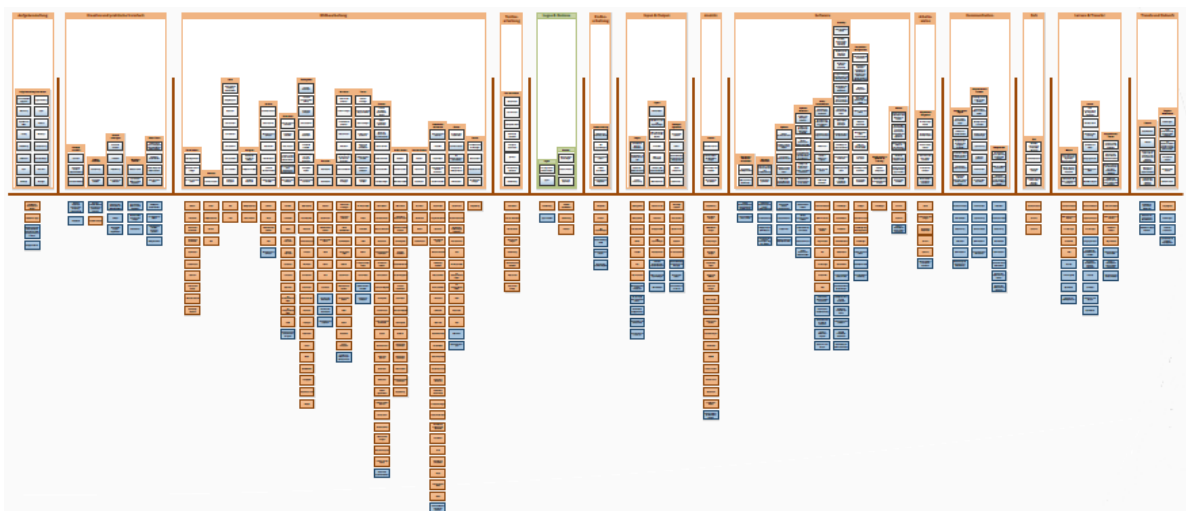


Abb.1: Mental Models Diagramm der Software GIMP nach Benning et al. (2012)

Dem werden die Funktionen der Software in Form eines Ähnlichkeitsdiagrammes (vgl. Abbildung 1) gegenübergestellt, sodass Bereiche aufgedeckt werden können, die der Nutzer sich bezüglich der Software zwar vorstellt, die aber bisher in dieser noch nicht umgesetzt sind. Zu diesen Gestaltungslücken (Gaps) können anschließend Lösungsvorschläge generiert werden, die in der weiteren Überarbeitung der Software berücksichtigt werden können. Wie genau diese Gaps zu gestalten sind lässt die Methode offen.

An dieser Stelle bietet sich die Verbindung mit der Methode der Image Schemata an. Durch die Methode können Komponenten einer Software nach sensumotorischem Vorwissen der Nutzer in Form von Bausteinen unbewusster mentaler Modelle gestaltet werden. Der Begriff "Image Schemata" steht für allgemeine, abstrakte Repräsentationen, welche durch wiederkehrende physische Interaktionen mit der Umwelt entwickelt und erweitert werden, wie beispielsweise das Gießen von Wasser in ein Glas. Das Glas entspräche in diesem Falle dem CONTAINER-Schema, da es einen Innenraum aufweist, welcher durch eine Grenze vom äußeren Bereich abgetrennt ist. Dieser Innenraum kann verschiedenen Zuständen entsprechen, also gefüllt oder leer sein. Die insgesamt mehr als 40 verschiedenen Image Schemata beschreiben objektbezogene Zustände, räumliche Dimensionen oder kraftdynamische Relationen (Hurtienne 2011). Sie eignen sich als Metavokabular, um sensumotorisches Vorwissen auf technische Nutzungsoberflächen zu übertragen. Dabei liegt das Potential der Methode in der Möglichkeit zur image-schematischen Erweiterung, was das Strukturieren abstrakter Konzepte ermöglicht. Die sogenannten image-schematischen Metaphern ergeben sich aus wiederholter Korrelation zwischen physischen, sensumotorischen Erfahrungen und abstrakten, subjektiven Erfahrungen (Hurtienne 2011). In diesem Zusammenhang korreliert zum Beispiel der Füllstand eines Wasserglases mit der Menge an Wasser, die sich darin befindet. Die Erfahrung, dass der Wasserpegel nach oben steigt, wenn man das Glas weiter auffüllt, kann durch die image-schematische Metapher "More is UP- Less is DOWN" benannt werden (Hurtienne 2011). Zur Einbringung der Nutzerperspektive in die Gestaltung einer Software, werden bei der Methode der Image Schemata sprachliche Äußerungen der Nutzer nach Image Schemata durchsucht werden. Aus dem Kontext, in dem diese geäußert wurden, werden nach dem Grundsatz "Quelldomäne ist Zieldomäne" (Hurtienne 2011, S. 49) image-schematische Metaphern formuliert, welche in der Gestaltung der Elemente einer Software umgesetzt werden können.

Praktische Umsetzung der Methodenkombination am Beispiel von GIMP

Im Rahmen einer studentischen Projektarbeit "Mental Models für GIMP" (TU Berlin 2012) wurde die Software GIMP, welche ihre Nutzer bei der professionellen Bildbearbeitung unterstützt, nach der Vorgehensweise der Mental Models Methode analysiert. Dafür wurden insgesamt acht Personen, die regelmäßig mit GIMP ar-

beiteten gemäß den sechs Regeln für Mental Models Interviews (Young 2008) zu verschiedenen, anwendungsbezogenen Aspekten der Software befragt. Die Transkripte, welche insgesamt 88 Seiten betrugen, wurden anschließend nach Atomic Tasks durchkämmt, nach Affinität gruppiert und mithilfe des Programmes Microsoft Visio 2010 zur Mental Map (vgl. Abbildung 1, Bereich oberhalb der Trennlinie) zusammengefügt wurde. Die Mental Map bildet die verschiedenen Gruppen der Atomic Tasks (wie z.B. Bildbearbeitung, kreative und praktische Vorarbeit, etc.) ab, die als mentale Räume und Türme bezeichnet werden. Im weiteren Vorgehen wurden den Komponenten der Mental Map die Funktionen der Software gegenübergestellt, welche die jeweiligen Elemente der Nutzerperspektive unterstützen und als Content Map bezeichnet werden (vgl. Abbildung 1, Bereich unterhalb der Trennlinie). Im Laufe der sogenannten Gap-Analyse wurden anschließend durch Gegenüberstellung der Mental Map und der Content Map Gestaltungslücken (Gaps) aufgedeckt, die nahezu über alle Türme der Mental Map verteilt waren. Dabei wurden zum Teil ganze Türme der Nutzerperspektive (kreative und praktische Vorarbeit, Kommunikation, etc.) nicht durch Funktionen des Programmes unterstützt. Für die gefundenen Gaps wurden abschließend Lösungsvorschläge generiert, welche darauf ausgelegt waren, die Gaps zu schließen, sodass die Software in ihren funktionalen Komponenten den Vorstellungen der Nutzer angepasst werden konnte.

Die Anwendung der Methode der Image Schemata bestand darin, zunächst aus den Transskripten der acht Interviews Image Schemata zu extrahieren. Dafür wurden die Transskripte Satz für Satz durchgelesen, wobei im Zusammenhang bestimmten Ausdrücken auf bestimmte Image Schemata geschlossen werden konnte. So wiesen zum Beispiel Worte wie "oben" oder "auf" auf das UP-DOWN-Schema hin. Zu den gefundenen Image Schemata wurden anschließend die image-schematischen Metaphern gebildet und in einem Sammeldokument den Bereichen (mentale Türme und Räume) der Mental Map zugewiesen. Dieses Vorgehen war sinnvoll, da so die Zählung der image-schematischen Metapher vereinfacht und diese schneller mit den jeweiligen Komponenten des Mental Model Diagramms in Verbindung gebracht werden konnten. Das Sammeldokument der image-schematischen Metaphern lieferte eine Grundlage, auf der zum einen den Lösungsvorschlägen Gestaltungshinweise zugeordnet werden konnte, zum anderen ein Abgleich der Nutzungsoberfläche von GIMP mit den image-schematischen Metaphern erfolgte. Der Abgleich wurde umgesetzt, indem den Elementen der Content Map die jeweiligen, image-schematischen Metaphern gegenübergestellt wurden und diese in ihrer Funktionsweise auf Konformität mit den image-schematischen Metaphern überprüft wurden. Dabei konnten Elemente gefunden werden, deren Funktionsweise von den Vorstellungen der Nutzer abwich. So sprachen einige der Nutzer zum Beispiel während der Interviews davon, in GIMP Teile des Bildes "aufzuhellen" bzw. "abzudunkeln", was dem UP-DOWN-Schema entsprach. Nach Überprüfung der Funktion "Helligkeit/ Kontrast" (vgl. Abbildung 2) konnte festge-

stellt werden, dass zur Regulierung der Helligkeit in GIMP derzeit ein horizontaler Schieberegler (vgl. Abbildung 2) zum Einsatz kommt. Um die Funktion besser an die Vorstellung der Nutzer anzupassen, könnte bei der weiteren Überarbeitung der Software ein vertikaler Schieberegler angedacht werden, sodass die Funktionsweise dem aus den Interviews extrahierten UP-DOWN-Schema entspricht.

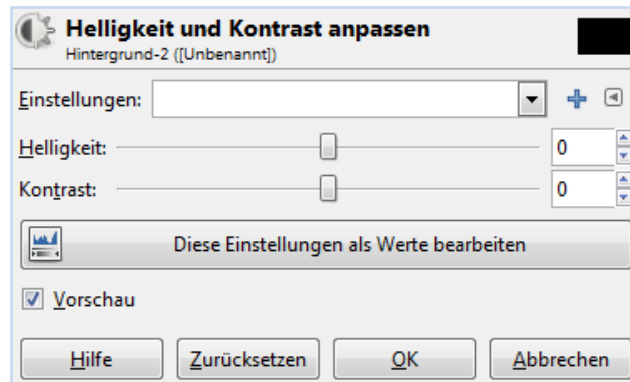


Abb.2: Dialogfenster zur Regulierung von Helligkeit und Kontrast

Nach dem Abgleich der Nutzungsoberfläche von GIMP mit den image-schematischen Metaphern wurden im weiteren Vorgehen exemplarisch Funktionen gestaltet, die zuvor als Gaps identifiziert worden waren. Dafür wurden zunächst zu jeder Funktion die image-schematischen Metaphern ausgewählt, die den Inhalt der zu gestaltenden Funktion am besten wiedergaben und für jede der Metaphern drei verschiedene Gestaltungsvorschläge entworfen. Die Gestaltungsvorschläge, welche nach Austausch mit anderen Image-Schemata-Experten am treffendsten gestaltet waren wurden in einen Low-Fidelity-Prototyp eingefügt und anhand von Nutzertests von potentiellen Nutzern bewertet. Dabei kam unter anderem die Methode des Lauten Denkens zum Einsatz, wobei die Nutzertests am Pluralistic Usability Walkthrough angelehnt waren. Die Nutzertests zeigten, dass die neugestalteten Funktionen im Großen und Ganzen in ihrer Gestaltung von den Nutzern verstanden wurden, allerdings zum Teil in ihrer Platzierung und Gestaltung nicht den Erwartungen der Nutzer entsprachen. So wurde zum Beispiel die Funktion "Bild hochladen" in der oberen Menüleiste des Bildfensters von GIMP platziert, sodass die hochzuladenden Bilder nach oben diese hineinzuziehen waren, was dem UP-Schema entsprach. Die Nutzer schienen durch diese Form der Interaktion überrascht, zum einen weil sie Drag & Drop-Interaktionen in Bildbearbeitungsprogrammen nicht gewohnt waren, zum anderen weil sie in der oberen Menüleiste des Bildfensters ein Dropdown-Menü erwartet hatten.

Diskussion

Die Kombination der Methoden Mental Models und Image Schemata eignet sich dafür, Software an die Vorstellungen der Nutzer anzupassen und auf Basis mentaler Repräsentationen zu gestalten und zu evaluieren. Dabei ist im Hinblick auf den Umfang der Methode abzuwägen, in welchem Falle eine Umsetzung der Methode

gerechtfertigt ist, da laut Young schon alleine das Durchkämmen von Interviewmaterial nach Atomic Tasks mehrere Tage andauern kann. Daneben werfen die Ergebnisse der Nutzertests die Frage auf, wie die Übertragung sensumotorischen Vorwissens mit spezifischem Vorwissen bezüglich Aufbau und Funktionsweise von Software vereint werden kann. In diesem Zusammenhang gilt es zu klären, wie vorgegangen werden muss, um konkurrierendes Vorwissen auf unterschiedlichen Stufen zu erfassen und auf die Gestaltung von Software zu übertragen.

Literatur

- Abras, C., Maloney-Krichmar, D. & Preece, J. (2004). User-Centered Design. In W. Bainbridge, *Encyclopedia of Human-Computer Interaction*. Thousand Oaks: Sage Publications.
- Benning, R., Chandra, A., Hillmann, C., Klose, E., Krause, T., Seelmann, A. & Zirk, A. (2012). *Mental Models für GIMP*. Unveröffentlichte Projektarbeit. Fachgebiet Mensch-Maschine-Systeme der Technischen Universität Berlin.
- Blessing, L. & Hurtienne, J. (2007). Design for Intuitive Use - Testing image schema theory for user interface design. In *ICED 2007 Paris, 16th International Conference on Engineering Design*, Proceedings of the conference (P_386, 1-12). Paris: Ecole Centrale.
- Hurtienne, J. (2011). Image Schemas and Design for Intuitive Use. Exploring New Guidance for User Interface Design. Dissertation, Technische Universität, Berlin.
- International Organization for Standardization (2010): ISO 9241-210: Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems.
- Mohs, C., Hurtienne, J., Israel, J. H., Naumann, A., Kindsmüller, M. C., Meyer, H. A. & Pohlmeier, A. (2006). IUUI- Intuitive Use of User Interfaces. In T. Bosenick, M. Hassenzahl, M. Müller-Prove & M. Peissner (Hrsg.). *Usability Professionals 2006* (S. 130- 133). Stuttgart: German UPA.
- Young, I. (2008). *Mental Models- Aligning Design Strategy with Human Behavior*. USA, New York: Rosenfeld.

Domain-Map Method for Communication Interface Optimization in Computer Assisted Orthognathic Surgery – a Pioneer Approach

Weichen Liu¹, Romy Lorenz², Nicolai Adolphs³, Bodo Hoffmeister⁴, Erwin Keeve⁵

Keywords: computer assisted orthognathic surgery, digital surgical splint generation, digital communication, user interface, domain map

Abstract

In computer assisted orthognathic surgery multilateral disciplines are required in a complex field. In the planning the malocclusion is surgically corrected (orthognathy) using software and surgical splints are made (Swennen, 2009; Metzger, 2008). Splint carries the relative position of upper and lower jaw in rigid body form. Such interdisciplinary workflow can be divided into steps which are concatenated into a process chain. In this work we accompanied and analyzed the most representative workflow variants. In ten clinical cases we evaluated the integrated digital process chain with parallel example of the ripe conventional variant (Profit, 2002). For optimization purpose the information- and the responsibility-transport were tracked and analyzed. During our work we developed a novel method, domain-map. With this pioneer tool we identified the problem pattern in the complex dependency mix-up. The working principle of domain-map was demonstrated on concrete examples.

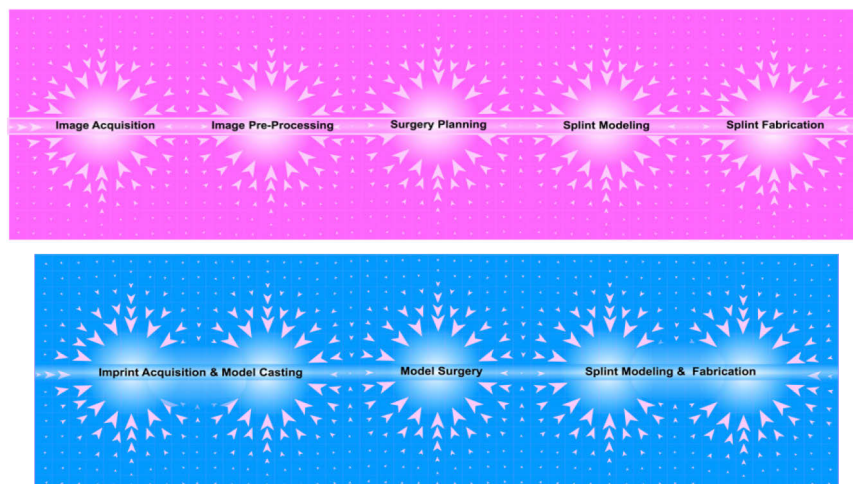


Fig.1: Domain-Map of the computer assisted planning (purple) and conventional planning (blue).

¹ Klinische Navigation und Robotik Charité Universitätsmedizin Berlin, Campus Virchow Klinikum

² Klinische Navigation und Robotik Charité Universitätsmedizin Berlin, Campus Virchow Klinikum

³ Klinik für MKG-Chirurgie Zentrum für rekonstruktive und plastisch-ästhetische Gesichtschirurgie

⁴ Klinik für MKG-Chirurgie Zentrum für rekonstruktive und plastisch-ästhetische Gesichtschirurgie

⁵ Fraunhofer Institute for Production Systems and Design Technology, Berlin, Germany

Methods

In Figure 1a domain map with fewer details shows the distribution of the involved competence. There are five major domains identified in the computer assisted planning, three (two of them are combines domains) in the conventional method. They are convergent in their own disciplines and span a one-dimensional field - a concatenated process chain. None of the domain is sharply bordered, but acts more like a gravitational field. In its kernel some ideals of this discipline are defined, they are theoretically formulated, but in the practice these are hard be realized. Attracted to the kernel are the affiliated with corresponding knowledge, skills and responsibility.

In Figure 2a detailed domain-map of the first involved competence - image acquisition (D1) with subdomain-division (white clouds) and responsibility-backtracking (threads with arrows) is demonstrated. Here the ideals of this domain are listed in the kernel. High resolution and low artifacts of the acquired image are obvious the most important goals among others. The pink threads (grey in black-white printing) are subdomain-connections based on the involved components in the imaging device. The backtracking unveils who is the responsible for pursuing certain ideal or rather who is to blame for occurrence of certain problem. The black thread visualized the command hierarchy (William, 2008) in the clinical workflow. As shown in the example domains and sub-domains have interfaces, they can be networked as knots. Information and responsibility can be transformed and transported among them. The domain-map method X-rayed the apparent interactivity in the multilateral workflow complex and showed the inner image of the essentials. We used this novel method to identify the most significant impacting factors and their adjustability in the process.

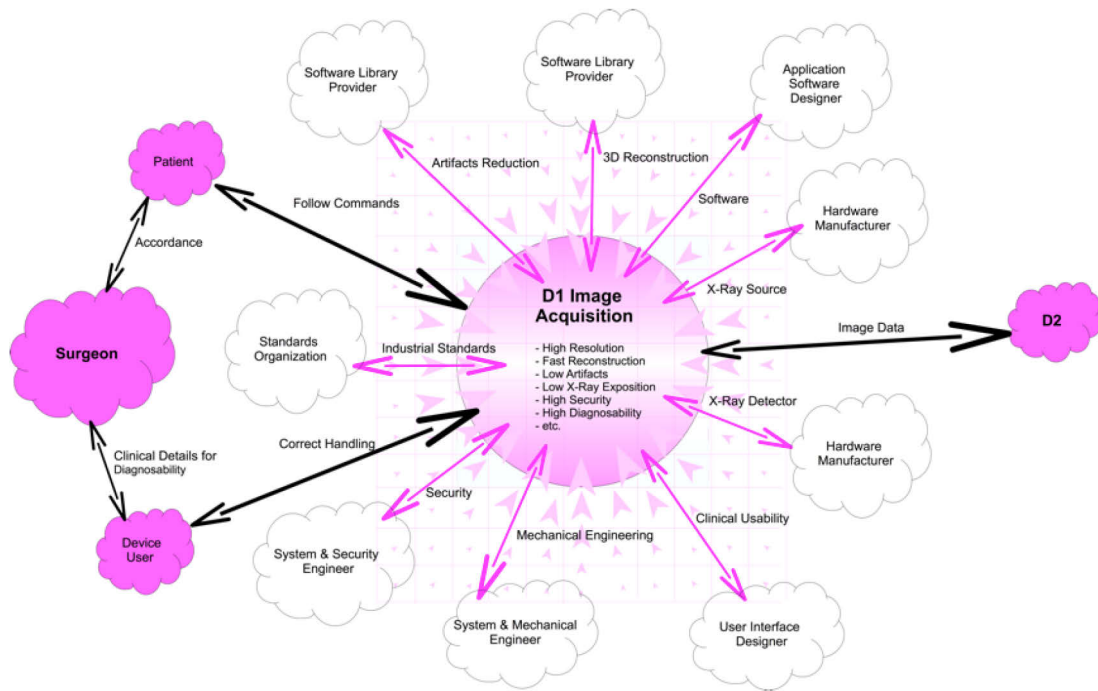


Fig.2: Detailed domain-map for D1 image acquisition.

We see in Figure 2 that the surgeon asks the patient to sign the accordance and commands an assistant to acquire the digital patient data using an imaging device (CT, CBCT, Surface-Scanner etc.). Since the surgeon has no essential competence in D1, he is dependent of the reliability of the imaging devices and has to be sure that his command is clearly formulated, understood and carried out. The reliability of such longtime proved and optimized modality as CT is high. So here the responsibility goes back to two major factors: the surgeon and the imaging device.

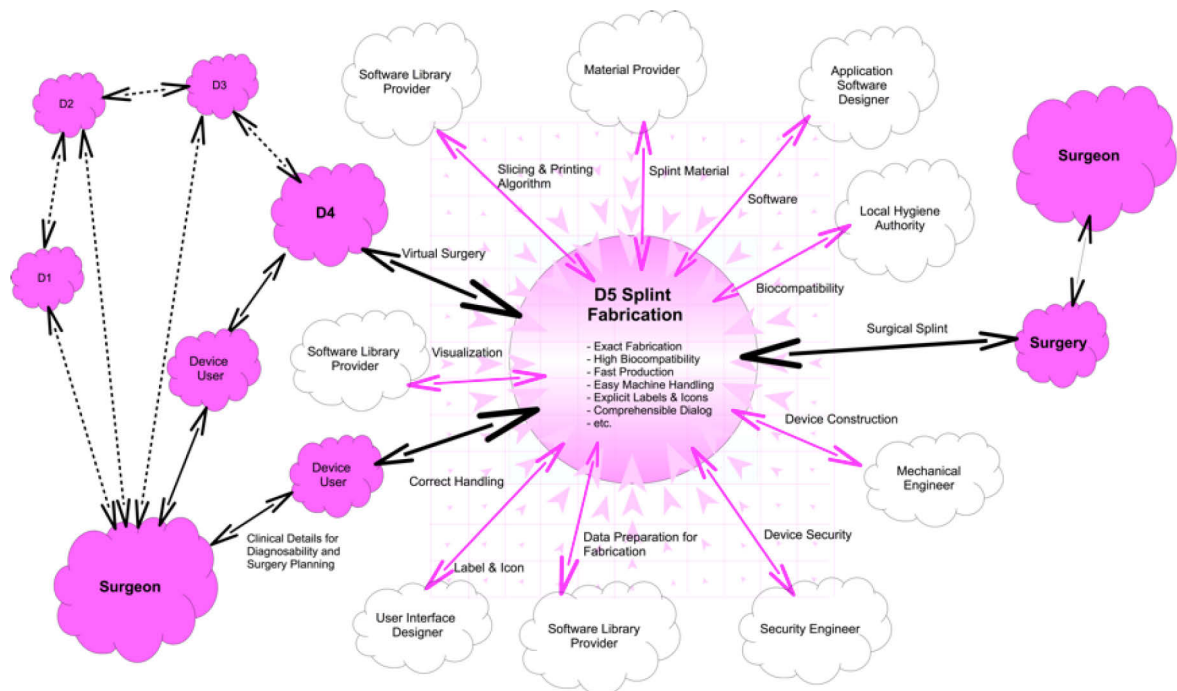


Fig.3: Detailed domain-map for D5 splint modeling.

Analogous to D1, D2, D3, D4 and D5 were analyzed. Only in D3 surgery planning the surgeon is in his own competence field. This step is the core of the whole process. D1 and D2 are its preparation, D4 and D5 are its finalization. Here the master plan was unfolded and the surgeon alone is responsible for the outcome. The dotted threads represent a simplified graph in the command hierarchy. At the end of the process chain a digital file with all splint geometry information will be imported and fabricated in D5 (Fig. 3). Here the mesh information will be processed (sliced) for the generative manufacturing. Only biocompatible (certified by local authority) materials can be used for the splint production since it has to be put into the mouth of the patient during the surgery. In this domain like in D1, the surgeon has no competence, he is dependent of the reliability of service provider or 3D-printer if there is one in the clinic. Usually the 3D printer will be delivered with a guaranteed production accuracy, which is normally under 100 micron for small sized shapes like splints. So in D5 the responsibility goes back to surgeon and the 3D-manufacturing machine.

Tab.1: Results of the clinical evaluation.

	Case	1	2	3	4	5	6	7	8	9	10
Konv.	Fitting	Yes	/	Yes	Yes	Yes	/	Yes	No	Yes	No
Digi.	Fitting	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes

In the clinical evaluation (in one year, Tab. 1) there were two times (case 5 and 8) in the digital workflow where the fitting of the splint failed. In the conventional workflow two cases failed, two cases blank since the dental technician is officially absent. In case 5, the modeling of splint failed, so its was in D4. The responsibility backtracking showed the failing-graph in black (Fig. 4). In this case the conventional splint was already finished and the surgeon commanded the device user to construct a conventional-like splint with local support for molars. But the support geometry was too big, the increased immersion resulted in higher fitting requirement. In the failing-graph there was a responsibility shift identified, where the device user was commanded by the surgeon to do a construction which he had no competence for. Analogous analyzes were made for case 8 and 10. Responsibility shift was also identified there.

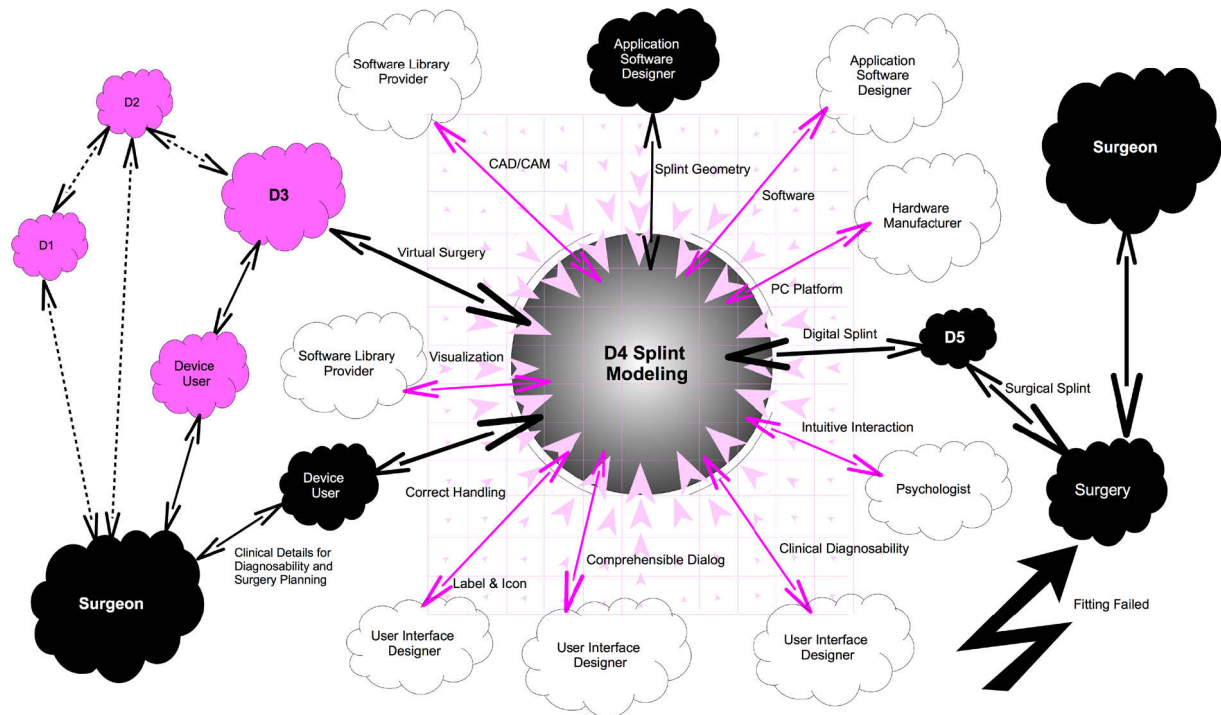


Fig.4: Responsibility backtracking of D4 for case 5.

Results & Discussion

Using the novel method domain-map the responsibility transport were analyzed. Responsibility shift was localized, where competence domain and user ID was mismatched. This was hidden in the process till it was too late to compensate. For optimization the communication interface has to be reformulated, which is more responsibility transparent. A warning dialog could help to make the commander more aware about the responsibility transport at critical steps. More responsibility shift sensitive communication protocol may help. This tool could also be used for analyzing and backtracking other complex interdisciplinary workflow.

Literature

- Metzger M., Hohlweg-Majert B. (2008). Manufacturing splints for orthognathic surgery using a three-dimensional printer. *Oral surgery, oral medicine, oral pathology, oral radiology, and endodontics*. 105(2), e1-7. doi: 10.1016/j.tripleo.2007.07.040
- Swennen, G., Mollemans W. (2009). Three-dimensional treatment planning of orthognathic surgery in the era of virtual imaging. *Journal of oral and maxillo-facial surgery*. 67(10), 2080-92. doi: 10.1016/j.joms.2009.06.007
- Proffit, W.R. (2002). *Contemporary Treatment of Dentofacial Deformity*. Oxford: Elsevier
- William J. B. (2008). *Management and the Arts* (4th ed.). Taylor & Francis.

Vertrauen in Fahrerassistenzsysteme – Gestaltung von Fahrerassistenz und Vertrauenskalibrierung

Alexander Maus¹ und Dr. Mirko Junge²

Schlüsselwörter: Fahrerassistenz, ADAS, trust in automation, ISO 26262, G&R

Zusammenfassung

Im Wesentlichen ist die Entwicklung von Assistenzsystemen durch die Maximierung des Komforts für die Insassen motiviert, wobei durch einige Systeme prinzipiell auch ein Zugewinn an Sicherheit möglich scheint. Es besteht nun die Gefahr, dass als Komfortsysteme entwickelte und abgesicherte Assistenzsysteme von Nutzern als Sicherheitssysteme wahrgenommen werden und als solche benutzt werden. Daher wird vorgeschlagen, das (erwartete) Nutzungsverhalten von Fahrerassistenzsystemen bei der Auslegung dieser zu berücksichtigen. Es wird ein Ansatz skizziert um vorhersehbaren Fehlgebrauch im Rahmen der Gefahren- und Risikoanalyse bereits bei der Entwicklung von Fahrerassistenzsystemen Rechnung zu tragen. Hierfür wird das Vertrauen des Nutzers in das jeweilige Assistenzsystem mittels einer noch im Entwicklungsstatus befindlichen Metrik erfasst und in einen Zusammenhang mit der Vertrauenswürdigkeit des betreffenden Systems ausgedrückt über das ASIL gestellt.

Abstract

Nowadays, Advanced Driver Assistance Systems (such as the parking assistant) are likely to be found in vehicles of all classes. Such systems are developed in accordance with the objective to maximize comfort. Some of them may also increase safety. Due to users potential misperception of actual capabilities of comfort features, risk for foreseeable misuse may rise. In order to face this problem within hazard analysis and risk assessment, an approach is suggested linking users trust in automation with the trustworthiness of respective ADAS. It is assumed, that miscalibrated trust may lead to improper usage of ADAS. The ASIL is suggested as a measure for trustworthiness in this context. A self-deployed psychometric instrument which is still under development will be used as a measure for user trust.

Einleitung

Im Jahr 2012 verloren insgesamt 3600 Personen im Straßenverkehr ihr Leben (Statistisches Bundesamt, 2013). Erklärtes Ziel ist es, die Zahl im Straßenverkehr Verletzter und Getöteter und somit die damit verbundenen persönlichen- und volkswirtschaftlichen Kosten so weit wie möglich zu reduzieren. Da ein überwiegender

¹ Automotive Safety Technologies GmbH (I/AST-5), Gaimersheim

² Volkswagen Aktiengesellschaft, Konzernforschung / Forschung Elektronik und Fahrzeug (K-EFFS/G), Wolfsburg

Anteil der Unfälle in der primären Unfallursache dem Menschen zugeschrieben werden kann (vgl. Mirwaldt, P., Bartels A., To, T.-B. & Pascheka, P. (2012)), ergibt sich hier ein möglicher Ansatzpunkt zur Erhöhung der Verkehrssicherheit für alle Beteiligten durch Fahrerassistenzsysteme. Dies wird von Fahrzeugkäufern ähnlich wahrgenommen (Robert Bosch GmbH Chassis Systems Control, 2013).

Automation

So unterschiedlich die am Markt befindlichen Systeme zur Maximierung von Komfort bzw. Sicherheit hinsichtlich der konkreten Funktion und natürlich der technologischen Realisierung sein mögen, eint sie doch das Grundprinzip der Automation. Für den Sicherheitsaspekt bedeutet dies, dass durch Automation die potentiell fehlerhafte Komponente Mensch im Mensch-Maschine-System im Bereich der Informationsaufnahme, Informationsverarbeitung, Entscheidung und Handlung aufgabenbezogen technologisch unterstützt oder gar durch möglichst weniger fehleranfällige technische Systeme ersetzt werden kann (Parasuraman, Sheridan & Wickens, 2000).

Im Weiteren wird folgende Definition von Automation verwendet, welche im Bereich der Fahrzeugtechnik wohl am ehesten mit dem Begriff des fortschrittlichen oder modernen Fahrerassistenzsystems (auch Advanced Driver Assistance Systems, ADAS) zu vergleichen ist. Automation is “[...] any sensing, detection, informationprocessing, decisionmaking, or control action that could be performed by humans but is actually performed by machine” (Moray, Inagaki & Itoh, 2000, S. 44).

Durch die Wahl dieser Definition werden automatisierte Systeme, die Aufgaben erfüllen, welche prinzipiell so nicht von einem Menschen erfüllt werden können (wie beispielsweise die Airbag-Steuerung) hier bewusst von der Betrachtung ausgeschlossen. Im Zentrum der Betrachtung stehen vielmehr ADAS, welche den Nutzer auf einem oder mehreren Ebenen menschlicher Informationsverarbeitung zu einem gewissen Grad in einer Aufgabe unterstützen, die sonst auch vom Fahrer durchgeführt werden könnte – was für die Auslösung eines oder mehrerer kaskadierter Airbagmodule in Sekundenbruchteilen kaum zu trifft.

Durch die beobachtbare Tendenz der Funktionsallokation vom Mensch hin zur Maschine ergeben sich jedoch insbesondere hinsichtlich der Interaktion des Nutzers mit dem System neue Herausforderungen. Die veränderte Aufgabenverteilung im Mensch-Maschine-System bedeutet für den Nutzer sowie für den Entwickler der Automation eine Anpassung an die sich verändernde Aufgabenstellung. Gelingt diese Anpassung auf einer der beiden Seiten nicht, kommt es bei der Interaktion zu Problemen (Sarter, Woods & Billings, 1997). So kann es beispielsweise durch die veränderte Interaktion und somit das veränderte Verhalten des Mensch-Maschine-Systems zu neuen Arten von Fehlern kommen, anstatt der angestrebten Vermeidung von Fehlern.

Parasuraman & Riley (1997) schlagen vor, Interaktion mit Automation als use, misuse oder disuse aufzufassen. Diese Begriffe werden hier wie folgt verwendet:

- use: Bestimmungsgemäße Verwendung eines ADAS,
- misuse: Nutzung eines ADAS außerhalb der Systemgrenzen, z.B. Nebelrasen mit eingeschaltetem ACC (vergleichbar mit dem Begriff des vorhersehbaren Fehlgebrauchs aus dem Bereich der funktionalen Sicherheit),
- disuse: Nichtbenutzung eines (potentiell hilfreichen) ADAS, z.B. Abschalten der Müdigkeitserkennung.

Um nun eine Aussage darüber treffen zu können, ob ein ADAS außerhalb der Systemgrenzen benutzt wird, muss ein Zusammenhang zwischen den objektiven Fähigkeiten der Automation und der Nutzungsintention bzw. dem zu erwartenden Nutzungsverhalten hergestellt werden. Ziel ist es hierbei, das Problem des vorhersehbaren Fehlgebrauchs aus einer nutzerzentrierten Perspektive zu beleuchten um die Gefahren- und Risikoanalyse um diesen Blickwinkel zu erweitern. Sollen bei der Auslegung von ADAS vorhersehbarer Fehlgebrauch wie beispielsweise das erwähnte Nebelrasen berücksichtigt werden, bedarf es einem Ansatz wie dem hier vorgeschlagenen der das (erwartete) Nutzungsverhalten in einen Zusammenhang mit den gegebenen Systemgrenzen stellt.

Vertrauen und Vertrauenswürdigkeit

Der hier vorgeschlagene Ansatz baut auf die Arbeit von Lee & See (2004) bzw. Lee & Moray (1994) auf – im speziellen auf das Konzept des kalibrierten Vertrauens. Es wird ein funktionaler Zusammenhang zwischen der Vertrauenswürdigkeit der Automation auf der einen Seite und dem Vertrauen des Nutzers in die Automation auf der anderen Seite hergestellt. Hierbei wird Nutzungsverhalten über die Nutzungsintention als Konsequenz der Einstellung des Nutzers gegenüber dem System – namentlich Vertrauen – modelliert.

Diese Herangehensweise Verhalten über Einstellungen vorherzusagen ist vergleichbar mit aus der Theory of Reasoned Action (TRA) nach Fishbein & Ajzen (1975) abgeleiteten Modellen zur Technologieakzeptanz (z.B. Technology Acceptance Model (TAM1) (Davis, 1989), TAM2 (Venkatesh & Davis, 2000) oder Unified theory of acceptance and use of technology (UTAUT) (Venkatesh, Morris, Davis & Davis, 2003)).

Um diesen Ansatz der indirekten Abschätzung von Nutzungsverhalten nun auf die vorliegende Fragestellung sinnstiftend anwenden zu können bedarf es einer validen Operationalisierung der (1) Vertrauenswürdigkeit des ADAS sowie (2) des Vertrauens das seitens des Nutzers in das System gesetzt wird.

Zur **Operationalisierung der Vertrauenswürdigkeit** eines ADAS respektive einer zu bewertenden Funktion bedient sich das hier vorgeschlagene Modell im Bereich der funktionalen Sicherheit. Die ISO 26262 formuliert Anforderungen auf

der Prozess-, System-, Hardware- und Softwareebene die es bei Entwicklung sicherheitskritischer Systeme im Automobilbau zwecks der Erreichung eines definierten Sicherheitsziels seit August 2011 zu erfüllen gilt.

Für den konkreten Anwendungsfall ist hierbei insbesondere der dritte Teil der ISO 26262 – welcher sich mit der Konzeptphase beschäftigt – relevant: “The objective of the hazard analysis and risk assessment is to identify and to categorise the hazards that malfunctions in the item can trigger and to formulate the safety goals related to the prevention or mitigation of the hazardous events, in order to avoid unreasonable risk.” (ISO, 2011, S. 6)

Im Rahmen der Gefahren- und Risikoanalyse (G&R) werden systematisch Versagenssituationen untersucht und dem jeweiligen Betrachtungsgegenstand (item) das ASIL (Automotive Security Integrity Level) zugeordnet. Grundlage für die Vergabe des ASIL sind:

- Exposition: Eintrittshäufigkeit der vorausgesetzten Situation,
- Controllability: Vermeidbarkeit des Versagens nach Eintritt der vorausgesetzten Situation,
- Severity: medizinischer Schaden.

Dabei gilt, je höher das vergebene ASIL (D ist das höchste Level), desto höher sind die Anforderungen an System sowie Entwicklungsprozess.

Die für das Erreichen des Sicherheitsziels notwendige Sicherheit der Implementierung ausgedrückt durch das ASIL wird nun als Operationalisierung für die Vertrauenswürdigkeit des zu betrachtenden ADAS bzw. der Funktion herangezogen. Es wird angenommen, dass ein Maß welches auszudrücken Vermag welches Sicherheitslevel bei der Entwicklung erreicht werden soll auch eine Aussage darüber zu lässt, wie viel Vertrauen dem entsprechenden System entgegengebracht werden darf.

Die **Operationalisierung von Vertrauen in Automation** bedarf vorab einer Begriffsklärung. Je nach wissenschaftlichem Kontext gibt es die unterschiedlichsten Möglichkeiten Vertrauen zu definieren. Die Arbeit von Lee & See (2004) liefert hier einen umfassenden Überblick und startet in der Betrachtung von Vertrauen in Automation bei der folgenden grundlegenden Definition: Trust is “[...] the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability [...]” (Lee & See, 2004, S. 54). Demnach ist Vertrauen als eine Heuristik zu Verstehen, die durch die Reduzierung von Komplexität handlungsleitend wirkt.

Lee & See (2004) weisen darauf hin, dass dieser Einfluss von Vertrauen auf das gezeigte Verhalten insbesondere in Situationen in denen eine ausführliche Evaluation möglicher Handlungsoptionen nicht geleistet werden kann zur Geltung kommt. Diese Begrenzung des sinnvollen Geltungsbereichs des mittelbaren Einflusses von Vertrauen auf das Verhalten ist für den Anwendungsfall der Fahrerass-

sistenz nicht von Nachteil. ADAS als Mittel zur Erhöhung der Verkehrssicherheit adressieren gerade diese Situationen in denen schnelles Handeln von Nöten ist und somit kein Raum gelassen wird für eine wohl überlegte analytische Gegenüberstellung der möglichen Handlungsoptionen durch den Fahrer.

Abhängig von der wissenschaftlichen Domäne werden nun die unterschiedlichsten Instrumente vorgeschlagen um Vertrauen gegenüber Personen, Organisationen, Marken oder eben technischen Systemen zu messen. Im Human Factors Bereich ist die bedeutende Rolle von Vertrauen in Automation gemeinhin anerkannt (Parasuraman, Sheridan & Wickens, 2008), jedoch sind etablierte Messmethoden rar gesät (z.B. Jian, Bisantz, Drury & Llinas (1998) oder Madsen & Gregor (2000)).

In Ermangelung einer etablierten Metrik für Vertrauen im vorliegenden Kontext wird basierend auf den erwähnten Arbeiten ein eigenes Instrument entwickelt. Dies ist zum Zeitpunkt der Einreichung der vorliegenden Arbeit Gegenstand andauernder Forschung. Die Konstruktion, Validierung und Normierung dieses Instrumentes soll hier daher nicht näher behandelt werden. An dieser Stelle sei nur erwähnt, dass der Fokus auf den vom (potentiellen) Nutzer subjektiv wahrgenommenen Systemeigenschaften des ADAS liegt, da hier seitens des OEM Potentiale zur Gestaltung und Vertrauenskalibrierung liegen. Dabei werden die wahrgenommene (1) Nützlichkeit, (2) Verlässlichkeit, (3) Systemtransparenz und (4) Verbundenheit mit dem System als Determinanten für Vertrauen angenommen. Diese latenten Variablen werden durch manifeste Variablen in Form eines Fragebogens erfasst. Die Fragebogenitems erfassen dabei die Zustimmung der Probanden zu Aussagen wie z.B. „Im täglichen Gebrauch ist das System nützlich für mich“ auf sechs-stufigen Likert-Skalen. Idealerweise sollte das ADAS hierbei für den Probanden erfahrbar sein wobei an dieser Stelle andere Wege der Präsentation der Funktionalität des jeweiligen Systems nicht ausgeschlossen werden (z.B. Demonstration von Prototypen oder anderweitige Darstellung des ADAS).

Abbildung 1 fasst nun den beschriebenen Zusammenhang zwischen Vertrauen des Nutzers in das jeweilige Assistenzsystem und der Vertrauenswürdigkeit des betreffenden Systems ausgedrückt über das ASIL schematisch zusammen. Analog zu Lee & See (2004) wird hier ein linearer Zusammenhang zwischen der Vertrauenswürdigkeit und dem Vertrauen hergestellt. Da das ASIL ordinalskaliert vorliegt, ergeben sich Bereiche in denen das Vertrauen als kalibriert angenommen wird. Hier ist von einem use-Fall auszugehen. Die Bereiche in denen das Vertrauen des Nutzers über der Vertrauenswürdigkeit liegt (overtrust) werden als potentielle misuse-Fälle eingeordnet. Analog hierzu werden Fälle in denen das Vertrauen des Nutzers in Anbetracht der Vertrauenswürdigkeit des Systems zu niedrig ausfällt (distrust) als potentielle disuse-Fälle angesehen.

Abbildung 1 lässt sich grundsätzlich in zweierlei Richtung interpretieren. Erstens kann ausgehend von einem vergebenen ASIL (im Beispiel ASIL A) auf den Be-

reich des im Sinne der Kalibrierung angebrachten Vertrauens geschlossen werden. Zweitens lässt sich im Umkehrschluss aus der Evaluation eines Systems hinsichtlich des zu erwartenden Vertrauens auf das (adäquate) ASIL schließen. In dem gewählten Beispiel in Abbildung 1 lässt sich zumindest für das fiktive System X begründet annehmen, dass es zu Problemen durch übersteigertes Vertrauen und somit misuse-Fällen kommen kann, da es unter der Bedingung kalibrierten Vertrauens bei der Projektion auf die Abszisse in den Bereich für ein ASIL B und somit aus dem Bereich des in diesem Fall tatsächlich vergebenen ASIL A fällt. Der Nutzer bringt dem System mehr Vertrauen entgegen als dieses einzulösen vermag.

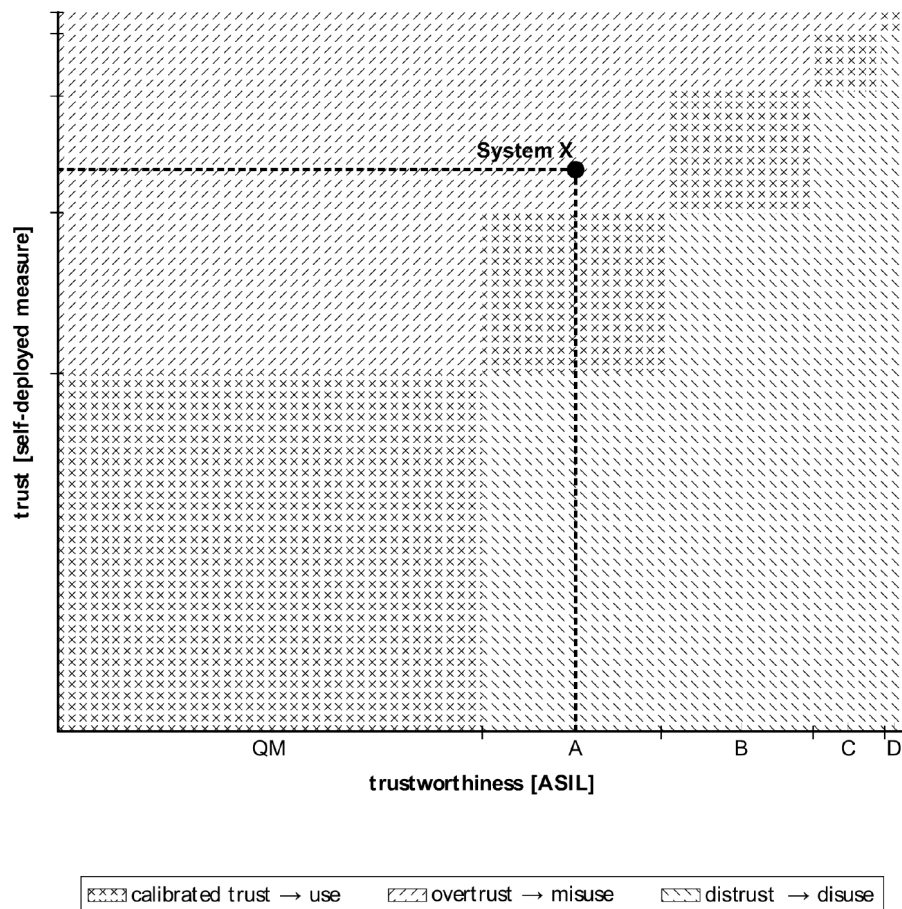


Abb.1: Exemplarische Darstellung des Zusammenhangs von Vertrauenswürdigkeit eines ADAS und der vom (potentiellen) Nutzer wahrgenommenen Vertrauenswürdigkeit.

Fazit

Aus dieser Betrachtungsweise ergibt sich ein Ansatzpunkt um im Rahmen der Gefahren- und Risikoanalyse vorhersehbaren Fehlgebrauch bereits während der Entwicklung von ADAS durch die Erfassung des Vertrauens (potentieller) Nutzer in das Assistenzsystem respektive die jeweilige Funktion zu berücksichtigen und dieses in einen funktionalen Zusammenhang mit dem jeweils vergebenen ASIL zu setzen. Für eine praktische Anwendung der skizzierten Methode bedarf es weitergehender Forschung hinsichtlich der Evaluation und Normierung des in Entwick-

lung befindlichen psychometrischen Messinstruments zur Erfassung des Vertrauens in Automation.

Literatur

- Robert Bosch GmbH Chassis Systems Control (2013). *Fahrerassistenzsysteme – wie viel Unterstützung wünschen deutsche Autofahrer?* Letzter Abruf 01.08.2013 http://www.bosch-presse.de/presseforum/download/de/7966ks-d_Anlage_Befragung_Fahrerassistenz.pdf
- Davis, F. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *Management Information Systems Quarterly*, 13(3), 319–339. doi:10.2307/249008
- Fishbein, M. & Ajzen, I. (1975). Belief, attitude, intention, and behavior: an introduction to theory and research. Reading, Ma, USA, Addison-Wesley Pub. Co.
- ISO (2011). ISO 26262-3:2011(E) Road vehicles – Functional safety – Part 3: Concept phase. Geneva, Switzerland, ISO copyright office.
- Jian, J.-Y., Bisantz, A. M., Drury, C. G. & Llinas, J. (1998). *Foundations for an empirically determined scale of trust in automated systems*. Letzter Abruf 01.08.2013 <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA395339>.
- Lee, J. D. & Moray, N. (1994). Trust, self-confidence, and operators' adaption to automation. *International Journal of Human-Computer Studies*, 40(1), 153–184. doi:10.1006/ijhc.1994.1007
- Lee, J. D. & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46 (1), 50–80. doi:10.1518/hfes.46.1.50_30392
- Madsen, M. & Gregor, S. (2000). *Measuring human-computer trust*. Vortrag gehalten auf Australasian Conference on Information Systems (ACIS), Brisbane. Letzter Abruf 01.08.2013 <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=233E62E02ADBD55A4253B5B493876D6F?doi=10.1.1.93.3874&rep=rep1&type=pdf>
- Mirwaldt, P., Bartels A., To, T.-B. & Pascheka, P. (2012). Gestaltung eines Notfallassistenzsystems bei medizinisch bedingter Fahrunfähigkeit. Vortrag gehalten auf 5. Tagung Fahrerassistenz, München. Letzter Abruf 07.08.2013 http://www.ftm.mw.tum.de/uploads/media/25_Mirwaldt.pdf
- Moray, N., Inagaki, T. & Itoh, M. (2000). Adaptive automation, trust, and self confidence in fault management of time-critical tasks. *Journal of Experimental Psychology*, 6(1), 44–58. doi:10.1037/1076-898X.6.1.44
- Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(2), 230–253.

- Parasuraman, R., Sheridan, T. B. & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans*, 30(3), 286-297. doi:10.1109/3468.844354
- Parasuraman, R., Sheridan, T. B. & Wickens, C. D. (2008). Situation awareness, mental workload and trust in automation: viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, 2(2), 140–160. doi:10.1518/155534308X284417
- Sarter, N., Woods, D. D., & Billings, C. (1997). Automation surprises. In G. Salvendy (Hrsg.), *Handbook of human factors & ergonomics* (2nd ed.). New York, USA, Wiley.
- Statistisches Bundesamt (2013). *Polizeilich erfasste Unfälle*. Letzter Abruf 07.08.2013
<https://www.destatis.de/DE/ZahlenFakten/Wirtschaftsbereiche/TransportVerkehr/Verkehrsunfaelle/Tabellen/UnfaelleVerunglueckte.html>
- Venkatesh, V. & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2) , 186–204. doi:10.1287/mnsc.46.2.186.11926
- Venkatesh, V., Morris, M., Davis, G. & Davis, F. (2003). User Acceptance of Information Technology: Toward a Unified View. *Management Information Systems Quarterly*, 27(3), 425–478.

Untersuchung der mentalen Repräsentation von Energiemanagement bei der Flugzeugführung zur Entwicklung eines Pilotenassistenzsystems

Simon Müller¹, Dietrich Manzey⁵⁵, Anne Bleyer⁵⁵, Karolin Schreiter², Anja Voigt⁵⁶ und Robert Luckner⁵⁶

Schlüsselwörter: Flugsimulatorversuche, Energiemanagement, Fragebogenstudie, Pilotenassistenzsystem, Mentales Modell

Zusammenfassung

Um die mentale Repräsentation des Energiemanagements bei der manuellen Flugzeugführung zu untersuchen, wurde eine Simulatorkampagne mit $n = 12$ lizenzierten Verkehrspiloten durchgeführt. Ihre Aufgabe war es, künstlich generierte Höhen- und Geschwindigkeitsablagen auf dem Gleitpfad des Instrumentenlandesystems (ILS) im Endanflug zu korrigieren. Dabei wurde untersucht, ob und inwiefern sie von dem Prinzip des Energieaustauschs Gebrauch machen. Die Ergebnisse zeigen eine wenig einheitlich ausgeprägte Repräsentation des Wissens bezüglich des Energiemanagements. Des Weiteren war zu erkennen, dass nicht die vollen Möglichkeiten des Energieaustausches genutzt wurden. Vielmehr wurden damit lediglich kleine Korrekturen durchgeführt.

Einleitung

Die manuelle Steuerung der Längsbewegung eines Flugzeugs lässt sich als Management von Gesamtenergie, bestehend aus potentieller und kinetischer Energie, zusammenfassen. Die Änderung der Flughöhe und der Fluggeschwindigkeit stehen hierbei in einem besonderen Zusammenhang. Diese beiden Flugbewegungsparameter sind direkt miteinander gekoppelt und lassen sich nur gemeinsam durch Kombination von Schub und Höhenruder steuern. Dabei wirkt sich der Schub auf die Gesamtenergie des Flugzeugs und das Höhenruder hauptsächlich auf die Verteilung der Gesamtenergie auf potentielle Energie und kinetische Energie aus (Brockhaus, Alles, Luckner, 2011).

Um eine energieeffiziente Bahnführung zu erreichen, muss der Pilot einschätzen können, welchen Energiestatus das Flugzeug zum aktuellen Zeitpunkt besitzt und wie dieser im Verhältnis zum gewünschten zukünftigen Energiestatus steht. Zur Regelung des aktuellen Flugzustands muss er sowohl die Abweichungen der Gesamtenergie vom Sollwert, als auch die Abweichungen von potentieller und kinetischer Energie vom Sollwert kennen. Dazu ist neben der Veränderung der Gesamt-

¹ Technische Universität Berlin, Institut für Psychologie und Arbeitswissenschaft, Fachgebiet Arbeits-, Ingenieur- und Organisationspsychologie

² Technische Universität Berlin, Institut für Luft- und Raumfahrt, Fachgebiet Flugmechanik, Flugregelung und Aeroelastizität

energie auch von großem Interesse, ob und wie sich potentielle und kinetische Energie anpassen lassen. Erfahrene Piloten können dieses Wissen zu Energiemanagement geschickt ausnutzen, um gewisse Ablagen energieeffizient zu korrigieren, in dem sie potentielle Energie und kinetische Energie gegeneinander austauschen.

So ist es beispielsweise möglich einen Flugzustand, in dem sich das Flugzeug über der Sollflughöhe und unterhalb der Sollgeschwindigkeit befindet, einzig durch Änderung des Nickwinkels zu korrigieren. Dadurch beschleunigt das Flugzeug und verliert gleichzeitig an Höhe, ohne dass zusätzlich der Schub oder der Widerstand angepasst werden muss.

Diese Zusammenhänge zu erkennen und zu nutzen zählt zu den schwierigsten Aspekten des manuellen Flugs (Amelink, Mulder, Paassen & Flach, 2005). Deshalb wird davon ausgegangen, dass gerade diese Aspekte viel Potential bergen, um eine Optimierung bezüglich energieeffizienter und ökonomischer Bahnführung aber auch bezüglich einer Verringerung der Arbeitsbelastung der Piloten zu erreichen. Wird das Prinzip des Energieaustauschs angewandt, spart das nicht nur Treibstoff, sondern es verringert auch die Belastungen, die mit einem ständigen Kontrollieren und Anpassen der Schubhebelstellung verbunden ist. Dieser Nutzen sollte sich im manuellen Erfliegen von zukünftigen komplexen Flugbahnen sogar noch intensivieren.

Verschiedene Maßnahmen erscheinen geeignet, den Piloten bei der anspruchsvollen Aufgabe des Energiemanagements im manuellen Flug zu unterstützen. Möglich wäre etwa eine Vermittlung des notwendigen Wissens und entsprechender Fertigkeiten durch ein gezieltes Training, die Integration von spezifischen Energieanzeigen in bestehende Bildschirmanzeigen (z.B. Primary Flight Display, siehe Lambregts, Rademaker & Theunissen, 2008 und Amelink et al., 2005) oder die Implementierung von Assistenzsystemen, die eine automatische Hilfe für effektives Energiemanagement bereitstellen.

Eine effektive Konzeption und Umsetzung von Unterstützungsmaßnahmen setzt voraus, dass Erkenntnisse darüber vorliegen, welches Wissen Piloten in Hinblick auf das Energiemanagement haben und unter welchen Bedingungen dieses Wissen wie eingesetzt wird. Nur so kann eine Planung von Unterstützungsmaßnahmen pilotenzentriert und effektiv erfolgen. Die in diesem Beitrag beschriebene Studie ist Teil einer umfassenderen Versuchsreihe. Die Versuchsreihe entstand im Rahmen des DFG-Projekts "nxControl" zur Entwicklung eines Pilotenassistenzsystems für die Unterstützung der manuellen Steuerung von Flugzeugen in Längsrichtung (für eine ausführliche Projektbeschreibung siehe Schreiter, Müller & Luckner, 2013). Dabei wurden die folgenden Annahmen für den Landeanflug untersucht:

1. Piloten unterscheiden sich in Bezug auf ihr Wissen über Energiemanagement,
2. nur Piloten, die Wissen über Energiemanagement vorweisen, wenden das Prinzip des Energieaustauschs an,

3. nur wenn die Ablage gering bzw. die Gesamtsituation nicht sicherheitskritisch ist, wird das Prinzip des Energieaustauschs angewandt.

Im Folgenden wird lediglich der für die Fragestellungen nach der mentalen Repräsentation von Energiemanagement entscheidende Teilausschnitt der Simulatorkampagne im Detail beschrieben (für weiterführende Informationen zur gesamten Versuchsreihe siehe Müller, Schreiter & Manzey, 2013).

Methode

Stichprobe: An den Versuchen nahmen zwölf Verkehrspiloten (11 männlich, 1 weiblich) teil. Durchschnittlich gaben sie rund 10000 Stunden ($SD \approx 8000$) Flug Erfahrung an. Die Altersspanne der Piloten lag zwischen 27 und 67 Jahren ($M = 43$, $SD = 12,5$). Sechs der zwölf Teilnehmer waren als Kapitäne angestellt, wobei sich einer der Kapitäne seit zwei Jahren im Ruhestand befand. Die Piloten besaßen Typenzulassungen für A320, A330 und/oder A340. Sieben Piloten gaben an, aktuell für Kurzstrecke eingeteilt zu sein. Fünf der zwölf Piloten besaßen einen Ingenieurshintergrund.

Aufgabe und Versuchsumgebung: Aufgabe der Versuchsteilnehmer war es in acht vorgegebenen Szenarien künstlich generierte Ablagen während eines Anflugs auf dem Gleitpfad zu korrigieren. Im Anschluss an die Versuchsdurchführung waren die Versuchspersonen aufgefordert, einen Wissensfragebogen zum Thema Energiemanagement zu beantworten.

Die Versuche wurden im Festsitz-Forschungssimulator SEPHIR durchgeführt. Das simulierte Flugzeug ist eine VFW614-ATD, welche über ein Flugsteuersystem, ähnlich dem eines modernen Verkehrsflugzeugs, verfügt. Obwohl geringe Unterschiede vorhanden sind, ist der Simulator in Bezug auf Flugeigenschaften und Bedienung einem Airbus A320 sehr ähnlich (für eine detaillierte Beschreibung siehe Müller et al., 2013).

Versuchsplan: Um beobachten zu können, ob Piloten Ablagen mithilfe des Prinzips des Energieaustauschs korrigieren, wurde ein 2 (Gruppe) \times 4 (Szenario) Versuchsplan entwickelt. Der erste Faktor war ein between-subjects-Faktor, der über eine post-hoc-Unterteilung der Stichprobe in zwei Gruppen mit unterschiedlichem Wissen über Energiemanagement operationalisiert wurde. Die Einteilung geschah anhand der Ergebnisse des Wissenstests in einem Debriefing-Fragebogen (siehe unten). Durch die Auswertung der Antworten des Fragebogens wurde die Stichprobe in eine Gruppe von Piloten mit explizitem Wissen zu Energiemanagement aufgeteilt und eine weitere mit Piloten, die dieses Wissen gar nicht oder nur fehlerhaft repräsentiert hatten. Bei dem zweiten Faktor handelte es sich um einen Messwiederholungsfaktor, in Form von vier verschiedenen Ablageszenarien, die im Folgenden genauer beschrieben werden.

Innerhalb der Ablageszenarien sollten die Versuchsteilnehmer eine künstlich generierte Höhendifferenz zum 3°-ILS-Gleitpfad und gleichzeitig eine Geschwindigkeitsablage zur Landegeschwindigkeit (115 kt) korrigieren. Der Referenzstartpunkt der Szenarien wurde dabei auf 2000 ft über Grund festgelegt. Das Flugzeug war zu dem Zeitpunkt bereits für den Endanflug komplett konfiguriert. Die Szenarien in Tab.1 wurden so berechnet, dass die Korrektur der Ablagen einzig durch den Austausch von potentieller und kinetischer Energie möglich ist, ohne die Gesamtenergie beeinflussen zu müssen. Anders ausgedrückt, können die Piloten die Korrektur bei diesen Szenarien ohne Schubanpassung ausschließlich durch Verwendung des Höhenruders ausgleichen. Durch Auswertung der Kommandogrößen des Schubhebels kann so erkannt werden, ob die Piloten das Potential zum Energieaustausch erkennen und in diesen Situationen auch anwenden. Unterschiede zwischen den Szenarien wurden hinsichtlich der Größe der Ablagen und des mit den Ablagen verbundenen Risikos realisiert. Ablageszenario 1 besitzt mit einer Höhendifferenz von -200 ft und Geschwindigkeitsdifferenz von +18,2 kt die größten Ablagen. Ablageszenario 4 gelangt durch die Geschwindigkeitsdifferenz von -10 kt gerade an die Toleranzgrenze für die niedrigste wählbare Geschwindigkeit. Die Szenarien 2 und 3 wurden als abgeschwächte Varianten der Szenarien 1 und 4 konzipiert.

Tab.1: Auflistung der Szenarien mit generierten Höhen- und Geschwindigkeitsablagen.

	Höhenablage	Geschwindigkeitsablage
1	-200 ft	+18,2 kt
2	-100 ft	+9,5 kt
3	+50 ft	-5 kt
4	+97,3 ft	-10 kt

Die vier Ablageszenarien wurden von den Versuchspersonen im Wechsel mit vier weiteren Ablagekombinationen geflogen. In diesen zusätzlichen Ablageszenarien ist es nicht möglich die Korrektur ausschließlich durch Energieaustausch durchzuführen. Die Reihenfolge der Szenarien wurde variiert, um Lerneffekte ausschließen zu können. Die vier weiteren Ablageszenarien werden in dieser Arbeit nicht behandelt.

Für jede Versuchsperson und jedes Szenario wurden die aufgezeichneten Simulatordaten abgespeichert und aufbereitet. Um eine Aussage zur Güte und zum Umfang des jeweils eingesetzten Energieaustauschs zu machen, wurde die Verwendung des Schubhebels betrachtet. Die Schubhebelstellung wurde durch den Versuchsleiter vor Beginn der Simulation auf den optimalen Wert zur Korrektur der Ablage durch Energieaustausch gesetzt. Deshalb war jede Abweichung zu dieser Startschubhebelstellung, die durch Schubregulation der Piloten herbeigeführt wurde, als Indikator dafür aufzufassen, dass die Piloten bei der Ablagekorrektur nicht allein das Prinzip des Energieaustausches nutzen. Vielmehr war dies ein Hinweis dafür, dass sie aktiv Energie zu- bzw. abführten. Aus dem Verlauf der Da-

ten zur Schubhebelstellung wurde pro durchgeführtem Szenario als abhängige Variable die Wurzel des mittleren quadratischen Fehlers (RMSE) ermittelt. Je größer dieser Wert ist, desto stärker fallen die Abweichungen zur optimalen Schubeinstellung aus und desto stärker hat ein Pilot die Schubregelung zur Ablagekorrektur eingesetzt. Je kleiner die RMSE-Werte desto besser und effektiver wird somit das Prinzip des Energiemanagements bei der Korrektur der Ablagen eingesetzt. Zur Prüfung der o.g. Annahmen wurden die so ermittelten Daten mit einer 2 (Gruppe) \times 4 (Szenario) Varianzanalyse mit Messwiederholungen über den zweiten Faktor (Szenario) analysiert.

Durchführungsdetails: Vor der Durchführung der Versuche wurde den Teilnehmer in einem Briefing die Versuchsaufgaben und der Simulator vorgestellt. In einem Probedurchlauf, vor der eigentlichen Durchführung des Versuchs mit Messaufzeichnung, konnten die Versuchsteilnehmer sich an die Eigenschaften des Simulators und des simulierten Flugzeuges gewöhnen. Dazu wurde ein ILS-Anflug mit Landung auf den Flughafen Frankfurt durchgeführt. Anschließend folgte die Versuchsdurchführung des ebenfalls durchgelaufenen Versuchs, der in Müller et al. (2013) detailliert beschrieben ist.

Im Anschluss an die Aufgaben aus dem ersten Versuch sollten die Piloten acht kurze Szenarien, darunter die vier hier besonders relevanten, abfliegen. Dazu wurde jeweils die Simulation eingefroren und auf die festgelegten Ablagen zurückgesetzt. Der Schubhebel wurde durch den Versuchsleiter jeweils auf den Wert für einen stationären Sinkflug mit 3°-Gleitwinkel zurückgestellt. Bevor die Simulation fortgeführt wurde, bekamen die Versuchsteilnehmer Zeit sich mit der gegebenen Situation vertraut zu machen und auf die Korrektur der Ablagen vorzubereiten. Aufgabe der Piloten war es anschließend bei aktiver Simulation die Ablagen wie im realen Linienflug spätestens bis zu einer Flughöhe von 1000 ft über Grund zu korrigieren.

Im Anschluss an die Versuchsdurchführung im Simulator wurde ein Debriefing-Fragebogen bearbeitet. Dieser war in Interviewform durch die Versuchsteilnehmer zu beantworten. Er dient dazu die mentale Repräsentation des Energieaustauschs zu untersuchen und das Verständnis der Piloten zu diesem Prinzip zu ermitteln. Der für die vorliegende Untersuchung relevante Teil des Fragebogens beinhaltet einen Wissenstest bestehend aus sieben Fragen zum Thema Energiemanagement.

Ergebnisse

Die post-hoc-Unterteilung der Stichprobe in zwei Wissensgruppen geschah anhand der Fragebogenantworten nach folgendem Schema. Die Versuchsteilnehmer, deren Antworten zur Beschreibung des Energiemanagements richtig sind, gehören der Wissensgruppe 1 ($n = 7$) an. Ist mindestens eine der Antworten falsch, wurde der Versuchsteilnehmer der Gruppe 0 ($n = 5$) zugeordnet.

Für eine Versuchsperson fehlten aufgrund defekter Datenaufzeichnung Messdaten zu Ablageszenario 2. Somit wurden alle Daten dieser Versuchsperson in folgenden statistischen Betrachtungen ausgeschlossen. Wissensgruppe 0 reduziert sich dadurch auf $n = 4$ Versuchspersonen.

In der Analyse der RMSE-Daten für die Schubhebelstellung wurde nur der Haupteffekt Szenario signifikant, $F(3, 27) = 10,049$; $p < 0,001$. Weder für den Haupteffekt Gruppe, noch für die Wechselwirkung Gruppe \times Szenario zeigten sich signifikante Effekte (beide $F < 2,0$).

Abb.1 zeigt die Mittelwerte der RMSE aller betrachteten Versuchspersonen für die Energieaustauschszenarien 1 bis 4. Es wird deutlich, dass in Szenarien 2 und 3 stärker auf das Prinzip des Energieaustauschs zur Beseitigung der Ablagen gesetzt wurde, als in den beiden Szenarien 1 und 4. Die Ablageszenarien 2 und 3 unterscheiden sich zu den Szenarien 1 und 4 dadurch, dass die Geschwindigkeits- und Höhendifferenzen geringer ausfallen (siehe Tab.1).

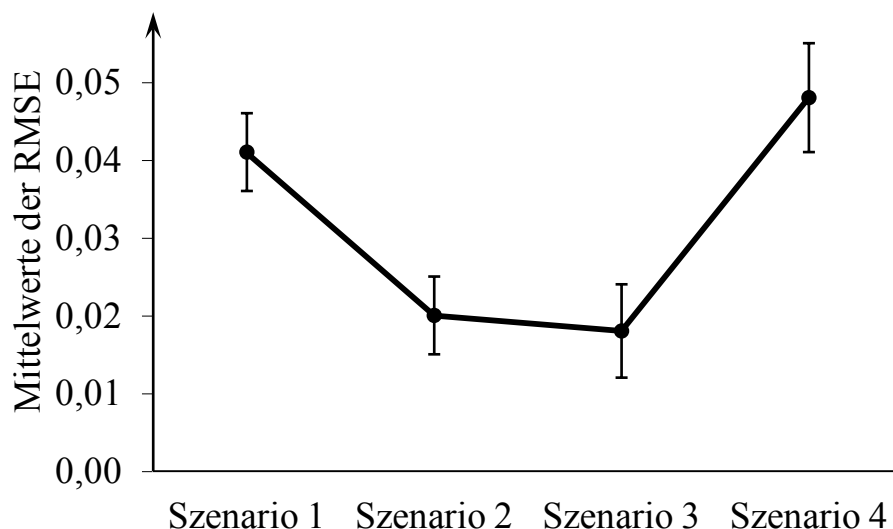


Abb.1: Diagramm der RMSE-Mittelwerte der Versuchspersonen mit Standardfehler über den Szenarien 1 bis 4.

Diskussion

Die Auswertung der Versuche zeigt, dass einige Piloten zwar ein Verständnis für Energiemanagement bei der Flugzeugführung besitzen, dieses Prinzip aber, unabhängig vom Wissen, nicht optimal eingesetzt wird. Die Annahme, nur Piloten, die Wissen bezüglich des Energiemanagements haben, wenden dieses auch an, ließ sich nicht bestätigen. Innerhalb der verschiedenen Ablageszenarien ließen sich jedoch signifikante Unterschiede beobachten, die die dritte Annahme bestätigen. Ablagen mit geringerem Korrekturbedarf wurden vermehrt mit dem Prinzip des Energieaustauschs ausgeglichen. Dagegen zeichneten sich die Szenarien mit größeren Ablagekorrekturen sowie Geschwindigkeitsablagen in sicherheitskritischer Toleranznähe durch stärkeren Einsatz des Schubhebels aus.

Daraus lässt sich schließen, dass auch Piloten mit explizitem Energiemanagementwissen selten das Prinzip des Energieaustauschs optimal anwenden. Insbesondere bei großen Ablagen besteht durch bessere Schnittstellengestaltung Potential zur Ressourcenoptimierung. Die Integration einer zusätzlichen Anzeige zur Darstellung des aktuellen und zukünftigen Energiezustandes würde es Piloten erleichtern die Möglichkeiten der Flugbahnführung besser einzuschätzen. In Kombination mit einer entsprechenden Schulung könnte so die manuelle Steuerung des Flugzeuges energetisch optimiert werden. Für Piloten wäre es damit möglich Korrekturen der Flugbahn nicht alleine durch Vorgaben der Höhen und Geschwindigkeitsänderung zu realisieren, sie können so direkt auf die Änderung der Energiezustände Einfluss nehmen. Beispielsweise wird dadurch ein möglicher Energieaustausch bei Ablagekorrekturen erkennbar oder Hinweise zur energetisch optimierten Schubwahl deutlich.

Energieanzeigen sind heutzutage kaum in der kommerziellen Luftfahrt vorhanden. Allerdings gibt es erste Konzepte die diesen Ansatz verfolgen (siehe Amelink et al., 2005 oder Lambregts et al., 2008). Für zukünftige Studien empfiehlt sich die Wirksamkeit dieser Konzepte zu untersuchen und die erwartete Verringerung der Arbeitsbelastung zu prüfen.

Abschließend sei noch darauf hingewiesen, dass die hier vorgestellten Ergebnisse nur mit Vorsicht verallgemeinert werden sollten, da die getestete Stichprobe vermutlich wenig repräsentativ für die Gesamtheit aller Verkehrspiloten ist. Dies begründet sich erstens in dem geringen Stichprobenumfang, der für diese Untersuchung zur Verfügung stand, und zweitens in dem unüblich hohen Anteil von Versuchsteilnehmern mit Ingenieurshintergrund.

Ebenfalls sollte nicht unerwähnt bleiben, dass die berechneten Ablageszenarien zwar theoretisch durch reinen Energieaustausch korrigierbar waren, sich aber bei unterschiedlichen Korrekturstrategien Verlustenergien ergeben konnten, die trotz Anwendung des Energieaustauschprinzips durch eine Schubanpassung korrigiert werden mussten. Die RMSE-Werte betreffender Versuchspersonen fallen daher zwar höher als im angenommenen Idealfall aus, sind jedoch im Vergleich zum Verzicht auf das Prinzip des Energieaustauschs deutlich kleiner.

Danksagung

Diese Arbeit wird gefördert durch die Deutsche Forschungsgemeinschaft (DFG), GZ: LU 1397/3-1, MA 3749/3-1.

Literatur

Amelink, M., Mulder, M., Paassen, M. M. (Rene) van, & Flach, J. (2005). Theoretical foundations for a total energy-based perspective flight-path display. *The*

International Journal of Aviation Psychology, 15(3), 205–231.
doi:10.1207/s15327108ijap1503

Brockhaus, R., Alles, W., Luckner, R. (2011). *Flugregelung* (3rd ed.). Berlin: Springer-Verlag.

Lambregts, T., Rademarker, R., & Theunissen, E. (2008). *A new ecological primary flight display concept*. Paper presented at Digital Avionics Systems Conference (pp. 4.A.1–1 – 4.A.1–20). St. Paul, MN.
doi:10.1109/DASC.2008.4702820

Müller, S., Schreiter, K., & Manzey, D. (2013). *Simulatorstudie zur Ermittlung mentaler Handlungsmodelle von Piloten beim Energiemanagement*. Paper presented at 62. Deutscher Luft- und Raumfahrtkongress. Stuttgart.

Schreiter, K., Müller, S., & Luckner, R. (2013). *nxControl: Konzept zur Vorgabe-
regelung für die Längsbeschleunigung des Flugzeugs*. Paper presented at 62. Deutscher Luft- und Raumfahrtkongress. Stuttgart.

Discount Usability-Maßnahmen als erste Schritte von KMU auf dem Weg zum Agile Usability Engineering

Ronny Reckin¹ und Stefan Brandenburg¹

Schlüsselwörter: Agile Usability Engineering, KMU, Ressourcen für das Usability Engineering, User Centered Design (UCD)

Zusammenfassung

Die vorliegende Arbeit gibt eine Übersicht darüber, wie Maßnahmen des Usability Engineering (UE) in kleine und mittelständische Unternehmen (KMU) eingebracht werden können. Dazu werden Schlüsselkriterien aus der Literatur extrahiert, deren Tauglichkeit an Praxisbeispielen evaluiert und Merkmale für ein Modell des agile Usability Engineering aufgestellt. Die Ergebnisse dieser theoretischen Herleitung legen nahe, dass UE für KMU vor allem schnell und erlebbar sein muss. Neuere Modelle des Usability Engineering unterstützen die Ergebnisse der vorliegenden Arbeit.

Einleitung

Seit den späten 1990'ern gewinnen agile Methoden der Softwareentwicklung (bspw. Scrum) zunehmend an Bedeutung im Softwareentwicklungsprozess (Boehm, 2006). Neben den Vorteilen dieser agilen Methoden (z. B. kurzen Entwicklungszyklen, frühe Zusammenarbeit mit den Käufern der Software) sind auch deutliche Nachteile dieser Herangehensweise bekannt. So geben sich agile Entwicklerteams oft mit der schnellsten Lösung zufrieden, anstatt sich um die langfristig erfolgreichste zu bemühen. Zumeist fehlen auch jegliche Anforderungsanalysen vor dem Entwicklungsbeginn, was die agilen Methoden in Reinform hauptsächlich für kleinere Softwareprojekte brauchbar macht (Boehm, 2006). Dementsprechend wird in der neueren Literatur insbesondere die Notwendigkeit von Methoden des Usability Engineering (UE) im agilen Entwicklungsprozess diskutiert (Gundelsweiler, Memmel, & Reiterer, 2004; Nebe, Düchting, & Zimmermann, Dirk, 2007). Hierbei wird beispielsweise die Bedeutung der Integration von Vorgehensweisen des UE in die agilen Methoden der Softwareentwicklung herausgestellt. Insbesondere für kleine und mittelständische Unternehmen (KMU) kann die Berücksichtigung der Gebrauchstauglichkeit ihrer Produkte ein wichtiger Erfolgsfaktor sein.

¹ Technische Universität Berlin, Institut für Psychologie und Arbeitswissenschaft

Förderinformation: Diese Arbeit entstand im Rahmen von „UseTree“ (www.usetree.de) und wurde durch das Bundesministerium für Wirtschaft und Technologie im Förderschwerpunkt Mittelstand Digital, Initiative Usability gefördert (Förderkennzeichen: 01MU12022A).

Erfolgskriterien für die Einführung von Usability Engineering bei agil entwickelnden KMU

In der Praxis ist diese Synthese von agilen Softwareentwicklungsmethoden und der Denkweise des Usability Engineering (UE) noch nicht erfolgt (Woywode, Mädche, Wallach, & Plach, 2011; Kessner, Dittrich, & Bär, 2011). Vielleicht die bedeutendste Ursache für diesen Unterschied von Theorie und Praxis liegt in dem enormen Zeit- und Kostendruck unter dem Softwarehersteller generell, und KMU im Besonderen, stehen (Kessner et al., 2011). Dieser Faktor spielt in den theoretischen Arbeiten zur Integration von agiler Softwareentwicklung und UE-Maßnahmen bisher keine bedeutsame Rolle. So erarbeiteten (Mommel, Brau, & Zimmermann, 2007) beispielsweise ein hybrides Vorgehensmodell, in dem agile Methoden und UE-Maßnahmen zum Agile Usability Engineering verschmolzen werden. Die Autoren gehen allerdings nicht explizit auf die (Mehr-) Kosten ein, die das „Plus“ an Usability im Vergleich zur herkömmlichen Anwendung der agilen Methode mit sich bringt. Auch in anderen Modellen und Studien zur Integration agiler Softwareentwicklung und UE-Maßnahmen (vgl. Tabelle 1) sind die notwendigen Investitionen in Einführung neuer Prozesse und Methoden und deren (dann zusätzliche) Anwendung bei der Softwareentwicklung nicht als wesentliches Erfolgskriterium berücksichtigt. Gerade bei kleinen Unternehmen sind die Möglichkeiten begrenzt, Usability-Experten zusätzlich einzusetzen und eine Verlängerung der Entwicklungszeit durch Änderung bewährter Abläufe und Kommunikation zu riskieren.

Tabelle 1. Weitere Modelle und Studien zur Integration von UE-Maßnahmen in den agilen Softwareentwicklungsprozess

Modelle	Studien
(Obendorf, Gibbert, Petersen, & Mommel, 2010)	(Chamberlain, Sharp, & Maiden, 2006)
(Beyer, Holtzblatt, & Baker, 2004)	(Ferreira, Sharp, & Robinson, 2012)
(Kane, 2003)	(Fox, Sillito, & Maurer, 2008)
(Obendorf & Naumann, 2007)	(Jia, Larusdottir, & Cajander, 2012)
(Mommel, Gundelsweiler, & Reiterer, 2007)	(Hussain, Slany, & Holzinger, 2009)
(Wolkerstorfer et al., 2008)	(Ungar & White, 2008)
(Sy & Miller, 2008)	(Sohaib & Khan, 2010)
(Düchting, Zimmermann, & Nebe, 2007)	(Silva da Silva, Martin, Maurer, & Silveira, 2011)

Der Einsatz leichtgewichtiger UE-Maßnahmen wird häufig als eines der wesentlichen Erfolgskriterien herausgestellt, um derartige Vorgehensweisen im agilen Softwareentwicklungsprozess zu verankern (z. B. Beyer et al., 2004; Mommel et al., 2007a). Sie sind schnell in der Durchführung, erfordern keine besondere Dokumentation und ihr Ergebnis steht direkt für den nächsten Überarbeitungszyklus der Software zur Verfügung. Die besondere Eignung von leichtgewichtigen UE-Maßnahmen für die regelmäßige Anwendung in der Praxis agiler Entwicklung

zeigt auch die Studie von Hussain et al. (2009). Die meisten der hier befragten Anwender agiler Methoden wendeten schnelle und unaufwendige Methoden an. Am häufigsten genannt wurden Low-Fidelity Prototyping, Conceptual Designs, Observational Studies of Users und Usability Expert Evaluations. Viele Befragte fanden zudem, dass die Integration der Maßnahmen ihren Entwicklungsprozess, ihr Team und die Produktqualität verbessert hätte. Die Ergebnisse von Jia et al. (2012) untermauern diese Aussagen. Sie fanden, dass von professionellen Scrum-Anwendern nur die relativ leichtgewichtigen UE- Maßnahmen regelmäßig (mehrmals wöchentlich oder monatlich) eingesetzt und als sehr nützlich bewertet wurden. Diese Maßnahmen umfassten Lo-Fi prototyping, Informal usability evaluation mit Nutzern, Scenarios, Digital prototyping, Personas, Usability goals. Formales Usability Testing wird hingegen von den Scrum-Anwendern sehr selten eingesetzt, weil es zeitlich nicht in die Projektarbeit passt (Larusdottir, 2011). Neben dem geringen Durchführungsaufwand wird eine enge Zusammenarbeit zwischen Entwicklern, User Interface Designern und Usability Experten und deren gute Kommunikation und Informationsteilung als wichtige Bedingung für ein erfolgreiches Agile Usability Engineering berichtet. Oft wird sogar die Bildung gemischter Entwicklerteams gefordert (Constantine, 2002; Gundelsweiler et al., 2004; Lárusdóttir et al., 2012; Nebe et al., 2007; Ungar & White, 2008). da Silva et al. (2011) arbeiteten in ihrem systematischen Literatur-Review bereits Schlüsselkriterien für die Integration von UE-Maßnahmen in die Agile Softwareentwicklung heraus. Darunter fielen: die Nutzung von 1) „Little Design Up Front“, 2) „Prototyping“, 3) „User Stories & User Testing“, 4) „Inspection evaluation“ und 5) „one sprint ahead“.

Evaluation der Kriterien in der Praxis

Die Evaluation der genannten Kriterien erfolgt anhand von Fallbeispielen. So beschreiben beispielsweise Stade, Reckin, Brandenburg und Thüning (2013) den Fall eines kleinen Unternehmens das die Software für Arztpraxen zum Patientenmanagement entwickelt. Zunächst wurden UE-Maßnahmen als zu teuer und zu umständlich benannt und aus dem Entwicklungsprozess ausgeschlossen. Nach einer kurzen Produktevaluation, die Prototyping und User Testing umfasste, änderte sich die Meinung des Unternehmens und die Bedeutung von UE-Maßnahmen wurde erkannt. Weiterhin berichten Kessner et al. (2011), dass die über 60 kleinen und mittelständischen Unternehmen, die in ihre Arbeit eingingen, hauptsächlich Expertenreviews (65%), User Stories und User Testing (22%) aber auch neue Prototypen und neue Design up fronts (11%) als Unterstützung für die Erhöhung der Gebrauchstauglichkeit ihrer Produkte benötigten. Demnach scheinen sich die Kriterien von da Silva et al. (2011) durchaus auch in der Praxis wiederzufinden.

Kriterien für ein Modell des Discount Usability Engineering

Soweit wurden Schlüsselkriterien für ein schnelles und erfolgreiches Verankern von UE-Maßnahmen in den agilen Softwareentwicklungsprozess dargestellt und

beispielhaft evaluiert. Wertvoller als diese Kriterien ist allerdings ein Modell des Discount Usability Engineering. Gegenüber den Kriterien hat ein Modell des Discount UE unter anderem den Vorteil, dass es angibt, wie sich die Berücksichtigung der Schlüsselkriterien auf unterschiedliche Parameter, wie beispielsweise die Bewertung der Nützlichkeit und die Häufigkeit der Anwendung von UE-Maßnahmen niederschlägt. Damit so ein Modell aufgestellt werden kann, müssen zunächst dessen Merkmale definiert werden. Aufgrund des Literaturüberblicks und der Schlüsselkriterien sollte ein Modell des Discount UE:

1. Eine niedrige Einstiegsschwelle schaffen, d.h. Betroffenheit herstellen und auf ein überschaubares Problem für eine einmalige Intervention fokussieren,
2. einen kurzen Zeitraum der Intervention vorschlagen (wenige Tage) und damit wenig Kosten produzieren und
3. betonen, dass die Entscheider (Management) deutlich den Mehrwert von UE-Maßnahmen erfahren sollten.

Ein Modell, dass diese Merkmale größtenteils aufweist wurde bereits von Stade et al. (2013) vorgeschlagen. Dahingehend durchläuft der Integrationsprozess von UE-Maßnahmen vier Schritte. In der Startphase wird der Problemraum eruiert und deutlich sichtbare Usability Probleme identifiziert. Die nachfolgende Pilotierungsphase umfasst die schnelle Bearbeitung eines Einzelprojekts mit deutlich sichtbarem Ergebnis. Die anschließende Expansionsphase erweitert den Aktionsradius auf weitere Usability-Probleme. Erst während der vierten, der Konsolidierungsphase, wird angestrebt die Anwendung von UE-Maßnahmen in die Produktentwicklungsprozesse zu integrieren.

Fazit und Ausblick

Die vorliegende Arbeit hatte zum Ziel herauszuarbeiten, welche Schlüsselkriterien für die erfolgreiche Einführung von UE-Maßnahmen in KMU gelten. Darüber hinaus wurden Merkmale benannt, die ein Modell des Discount UE erfüllen sollte. Bereits (Kane, 2003) fordert den Einsatz von Discount UE, wie es von Nielsen ausgearbeitet wurde (Nielsen, 1989), für die bessere Adressierung von Usability-Problemen bei agiler Entwicklung. Unser Ziel ist eine Anpassung des Discount UE auf die Rahmenbedingungen von KMU. Hierzu kann das Vorgehensmodell von Stade et al. (2013) genutzt werden. Es berücksichtigt die benannten Schlüsselkriterien und erfüllt die daraus abgeleiteten Merkmale. Eine umfassende Evaluation des Modells steht allerdings noch aus. Diese sollte Gegenstand weiterer Forschungsarbeiten sein.

Literatur

Beyer, H., Holtzblatt, K., & Baker, L. (2004). An Agile Customer-Centered Method: Rapid Contextual Design. In C. Zannier, H. Erdogmus, & L. Lindstrom

- (Eds.), *Extreme Programming and Agile Methods - XP/Agile Universe 2004* (S. 50–59). Springer Berlin Heidelberg.
- Boehm, B. (2006). *A View of 20th and 21st Century Software Engineering*. Proceedings der ICSE'06, Shanghai, China.
- Chamberlain, S., Sharp, H., & Maiden, N. (2006). Towards a Framework for Integrating Agile Development and User-Centred Design. In P. Abrahamsson, M. Marchesi, & G. Succi (Eds.), *Extreme Programming and Agile Processes in Software Engineering* (S. 143–153). Springer Berlin Heidelberg.
- Constantine, L. (2002). Process agility and software usability: Toward lightweight usage-centered design. *Information Age*, August/September 2002.
- Düchting, M., Zimmermann, D., & Nebe, K. (2007). Incorporating User Centered Requirement Engineering into Agile Software Development. In J. A. Jacko (Ed.), *Human-Computer Interaction. Interaction Design and Usability* (S. 58–67). Springer Berlin Heidelberg.
- Ferreira, J., Sharp, H., & Robinson, H. (2012). *Agile Development and User Experience Design Integration as an Ongoing Achievement in Practice*. In Agile Conference (AGILE), 2012 (S. 11–20). Presented at the Agile Conference (AGILE), 2012. doi:10.1109/Agile.2012.33
- Fox, D., Sillito, J., & Maurer, F. (2008). *Agile methods and User-Centered design: How these two methodologies are being*. Proceedings of the Agile.
- Gundelsweiler, F., Memmel, T., & Reiterer, H. (2004). Agile Usability Engineering. In R. Keil-Slawik, H. Selke, & G. Szwillus (Eds.), *Mensch & Computer 2004: Allgegenwärtige Interaktion*. München: Oldenbourg Verlag.
- Hussain, Z., Slany, W., & Holzinger, A. (2009). Current State of Agile User-Centered Design: A Survey. In A. Holzinger & K. Miesenberger (Eds.), *HCI and Usability for e-Inclusion* (S. 416–427). Springer Berlin Heidelberg.
- Jia, Y., Larusdottir, M. K., & Cajander, Å. (2012). The Usage of Usability Techniques in Scrum Projects. In M. Winckler, P. Forbrig, & R. Bernhaupt (Eds.), *Human-Centered Software Engineering* (S. 331–341). Springer Berlin Heidelberg.
- Kane, D. (2003). *Finding a place for discount usability engineering in agile development: throwing down the gauntlet*. In Agile Development Conference, 2003 (S. 40–46). Proceedings of the ADC 2003.
- Kessner, D., Dittrich, F., & Bär, N. (2011). Von der Wissenschaft in die Wirtschaft - Wissenstransfer in Sachen Usability. In H. Brau, A. Lehmann, K. Petrovic, & M. C. Schroeder (Eds.), *Usability Professionals 2011* (S. 236–241). Stuttgart, German UPA e.V.
- Larusdottir, M., Cajander, A., & Gulliksen, J. (2012). The big picture of UX is missing in scrum projects. In E. L.-C. Law, S. Abrahao, & A. P. O. S. Vermeeren (Eds.), *Proceedings of International Workshop on the Interplay*

- between User Experience Evaluation and Software Development* (S. 65). University of Leicester, UK: Audio Visual Services (AVS).
- Larusdottir, M. K. (2011). Usability Evaluation in Software Development Practice. In P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2011* (S. 430–433). Springer Berlin Heidelberg.
- Memmel, T., Brau, H., & Zimmermann, D. (2007a). Agile nutzerzentrierte Softwareentwicklung mit leichtgewichtigen Usability Methoden – Mythos oder strategischer Erfolgsfaktor? In K. Röse & H. Rau (Eds.), *Usability Professionals 2007* (S. 223–227). Stuttgart: Fraunhofer IRB Verlag.
- Memmel, T., Gundelsweiler, F., & Reiterer, H. (2007). *Agile human-centered software engineering*. In Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it - Volume 1 (S. 167–175). Swinton, UK: British Computer Society.
- Nebe, K., Düchting, M., & Zimmermann, Dirk. (2007). Integration von User Centred Design Aktivitäten in Agile Softwareentwicklung. In K. Röse & H. Rau (Eds.), *Usability Professionals 2007* (S. 223–227). Stuttgart: Fraunhofer IRB Verlag.
- Nielsen, Jakob. (1989). Usability Engineering at a Discount. In: Salvendy, G., and Smith, M. J. (Eds.), *Designing and Using Human-Computer Interfaces and Knowledge Based Systems*, Elsevier Science Publishers, Amsterdam. 394 – 401, 1989.
- Obendorf, H., Gibbert, R., Petersen, I., & Memmel, T. (2010). Agile UX – Wege zur agilen nutzerzentrierten Entwicklung. In H. Brau, S. Diefenbach, K. Göring, M. Peissner, & K. Petrovic (Eds.), *Usability Professionals 2010*. Fraunhofer Verlag.
- Obendorf, H., & Naumann, R. (2007). Extreme Gebrauchstauglichkeit: Agile Prozess- und Softwareentwicklung mit den Mitteln des Usability Engineering. In K. Röse & H. Rau (Eds.), *Usability Professionals 2007*. Stuttgart: Fraunhofer IRB Verlag.
- da Silva, T., Martin, A., Maurer, F., & Silveira, M. (2011). *User-Centered Design and Agile Methods: A Systematic Review* (S. 77–86). IEEE. doi:10.1109/AGILE.2011.24
- Sohaib, O., & Khan, K. (2010). *Integrating usability engineering and agile software development: A literature review*. In 2010 International Conference on Computer Design and Applications (ICCD) (Vol. 2, pp. V2–32–V2–38). Proceedings ICCDA. doi:10.1109/ICCD.2010.5540916
- Stade, M., Reckin, R., Brandenburg, S., & Thüring, M. (2013). Usability in KMU etablieren: Von schneller Problemlösung zu ressourcenorientiertem Usability Engineering. In: S. Boll, S. Maaß, & R. Malaka (Hrsg.). *Workshopband Mensch & Computer 2013*, S. 19 – 27, München: Oldenbourg Verlag.

- Sy, D., & Miller, L. (2008). *Optimizing agile user-centred design*. In CHI '08 Extended Abstracts on Human Factors in Computing Systems (S. 3897–3900). ACM.
- Ungar, J., & White, J. (2008). *Agile user centered design: Enter the Design Studio - A case study*. CHI08 extended abstracts on Human factors in computing systems, S. 2167–2178. doi:10.1145/1358628.1358650
- Wolkerstorfer, P., Tscheligi, M., Sefelin, R., Milchrahm, H., Hussain, Z., Lechner, M., & Shahzad, S. (2008). *Probing an agile usability process*. In CHI '08 Extended Abstracts on Human Factors in Computing Systems (S. 2151–2158). New York, NY, USA: ACM. doi:10.1145/1358628.1358648
- Woywode, M., Mädche, A., Wallach, D., & Plach, M. (2011). *Gebrauchstauglichkeit von Anwendungssoftware als Wettbewerbsfaktor für kleine und mittlere Unternehmen (KMU)* (Abschlussbericht). Institut für Mittelstandsforschung an der Universität Mannheim; Lehrstuhl für Wirtschaftsinformatik IV der Universität Mannheim; Fachhochschule Kaiserslautern; ERGOSIGN GmbH.

HLB und Web 2.0 – Anforderungen und Konzept für eine Expertenplattform

Ulrike Schmuntzsch¹ und Sarah-Christin Freytag¹

Schlüsselwörter: Expertenplattform, Web 2.0-Technologien, Onlinebefragung, Anforderungsanalyse, Hybride Leistungsbündel

Zusammenfassung

Dieser Werkstattbeitrag präsentiert erste qualitative Ergebnisse zweier Onlinebefragungen zu den Anforderungen an eine Expertenplattform in hybriden Leistungsbündeln (HLB). Die zu entwickelnde Expertenplattform soll Servicetechnikern ermöglichen, die bspw. einen komplexen Wartungsauftrag ausführen, Expertenwissen über Probleme und Lösungen abzurufen und selbst einzustellen. Die Anforderungen wurden getrennt für HLB-Experten und Servicetechniker erhoben, um die Besonderheiten in HLB und die speziellen Anforderungen technischer Kontexte herauszuarbeiten. Die Ergebnisse zeigen, dass die prototypische Realisierung einer *intelligenten Suche* sowie eines *Wikis* für HLB-Experten und Servicetechniker am sinnvollsten erscheint. Die HLB-Experten erachten zudem die Realisierung eines *Videoportals* und die Servicetechniker ein *Diskussionsforum* als sinnvoll. Hinsichtlich der Motivationswirkung verschiedener Aspekte, wie z.B. *Gamification*, zeigen sich Unterschiede zwischen den Gruppen. Die HLB-Experten heben die Wichtigkeit von materiellen Vergünstigungen hervor und bezweifeln den Effekt altruistischer Werte auf die Bereitschaft von Servicetechnikern zur aktiven Beteiligung. Die Servicetechniker hingegen betonen die motivierende Wirkung, Kollegen mit eigenen Beiträgen bei technischen Problemen zu helfen.

1. Einleitung

Führende Industrienationen wie bspw. Deutschland sehen sich seit geraumer Zeit mit einem wachsenden globalen Wettbewerb mit aufstrebenden Ländern wie z.B. China konfrontiert. Im Bereich der industriellen Produktion wird versucht dieser Entwicklung durch eine zunehmende Spezialisierung und Komplexität der angebotenen Produkte und Services zu begegnen. Ziel dessen ist es, ein Alleinstellungsmerkmal und somit einen Wettbewerbsvorteil zu sichern (Uhlmann, 2011). Im Kontext hybrider Leistungsbündel (HLB) sind die zu einem Industrieprodukt zusätzlich angebotenen Leistungen daher keine reinen Sachleistungen, sondern vielmehr komplexe Dienstleistungsmodulare (z.B. Instandhaltung und Produktschulungen), die die mit dem Produkt oder der Maschine untrennbar verbunden sind. Je nach Geschäftsmodell können diese dann entsprechend als Bündel angeboten werden (Meier, Uhlmann & Kortmann, 2005). Damit die entstandenen Industrieprodukte, trotz aller Komplexität, für Arbeitnehmer bedienbar bleiben, ist die Ent-

¹ Technische Universität Berlin, Institut für Psychologie und Arbeitswissenschaft, Fachgebiet Mensch-Maschine-Systeme

wicklung nutzerfreundlicher Unterstützungskonzepte unerlässlich, die die Maschine und den angekoppelte Service einbeziehen.

Vor diesem Hintergrund ist das Ziel des Projektes die umfassende Unterstützung eines im industriellen Kontext arbeitenden Servicetechnikers zur Vermeidung von Fehlhandlung und zur Optimierung von Arbeitsprozessen. Hierfür wird u.a. eine Expertenplattform konzipiert, die das Verwenden, Einstellen und Bewerten von Inhalten, wie z.B. von Handlungsanleitungen, im HLB ermöglicht. Im Rahmen dieses Werkstattbeitrages werden erste qualitative Ergebnisse zweier Onlinebefragungen zu den Anforderungen an eine Expertenplattform im HLB vorgestellt.

2. Theoretischer Hintergrund zur Wissensteilung mit Web 2.0

Wissen zu teilen gewinnt in einer Welt stetigen technischen Fortschritts und zunehmender Komplexität, die Spezialisierung nötig macht, immer mehr an Bedeutung. Web 2.0 bietet Nutzern ein breites Feld an Möglichkeiten, ohne Programmierkenntnisse eigene Inhalte zu erstellen und zu verbreiten. Dies stellt das Web 2.0 den früheren Phasen des *World Wide Webs* gegenüber, in denen Nutzer vertiefte Kenntnisse benötigten, um mit eigenen Inhalten partizipieren zu können (Li & Poon, 2011). Wissensteilung kann im Web 2.0-Kontext u.a. mittels Blogs, Wikis, aber auch über Videosharing-Plattformen, Foren und soziale Netzwerke stattfinden (Chirumalla, Bertoni & Johansson, 2013, Tudorache, Falconer, Noy, Nyulas, 2010). Unterstützung mittels Wissen wird daher zunehmend durch Gewährleistung von Zugang zu den relevanten Ressourcen definiert, wobei die jeweils optimale Form der Vermittlung kontextabhängig ist.

Für HLB bedeutet dies, die Bedürfnisse potentieller Nutzer im Vorfeld zu erfassen. Als erster Schritt muss hierbei die Form des zu teilenden Wissens ermittelt werden. Wissen lässt sich grob in prozedurale und deklarative Anteile unterteilen (Anderson, 1982). Die notwendigen Bewegungsschritte, um bspw. eine Wartungsaufgabe auszuführen, bilden hierbei die prozedurale Wissenskomponente ab, während das Wissen über die Teile, die gewartet werden müssen, deklarativ ist. Hier wird deutlich, dass die Expertenplattform für die Übermittlung beider Wissensarten geeignet sein muss.

Wikis gelten als Synonym für Wissensteilung und haben ihre Eignung für verschiedene Wissensarten erwiesen (Luoma & Okkonen, 2009). Blogs sind ebenfalls hilfreich für die Teilung von Wissen über Arbeitserlebnisse und Sicherheitsfragen (Li & Poon, 2011). Luoma & Okkonen (2009) stellen jedoch heraus, dass Blogs nur eingeschränkte Kollaborationsmöglichkeiten für Nutzer bieten. Dies ermöglichen hingegen Wikis und Diskussionsforen (Tudorache et al., 2010).

Zu beachten ist, dass nicht alle Wissenskomponenten über schriftlichen Austausch gleich gut vermittelt werden können. Daraus ergibt sich die Notwendigkeit des Einsatzes anderer Medien. Hierfür hat sich die Unterstützung durch Videosequenzen, in denen Wartungsprozeduren per Avatar demonstriert werden, als sinn-

voll erwiesen (Schmuntzsch, Sturm, Reichmuth & Rötting, 2012). Auch im Kontext von HLB ist Videosharing als Ansatz geeignet (Chirumalla et al., 2013).

Im zweiten Schritt muss die Motivation der Nutzer zur aktiven Teilnahme sichergestellt werden. Hierfür stehen neben der nutzerfreundlichen Gestaltung der Plattform weitere Möglichkeiten zur Förderung zur Verfügung. Ford & Staples (2005) schlagen vor, die wahrgenommene Nähe der Nutzer untereinander sowie deren altruistische Werte anzusprechen. Da potentielle Nutzer im HLB auch Servicetechniker konkurrierender Unternehmen sein können, ist im Vorfeld abzuklären, inwieweit diese sich kennen bzw. aus altruistischen Motiven Wissen teilen wollen und können. Der *Gamification*-Ansatz ist eine weitere Möglichkeit, Nutzermotivation und -bindung zu fördern (Deteding, Dixon, Khaled & Nacke, 2011, Hulsebosch, 2013). *Gamification* wird definiert als die Nutzung von Game-Design Elementen in Kontexten außerhalb von Spielen (Deterring et al., 2011). Dies können einzelne Komponenten wie *Badges*¹, *Leaderboards*² und *Quests*³ sein, jedoch auch der grundsätzlich spielorientierte Aufbau der Plattform. Im wirtschaftlichen Kontext wird *Gamification* oft mit Web 2.0-Elementen verbunden. Um die motivierende Wirkung zu entfalten, muss *Gamification* bereits in der Planungsphase an die Zielgruppe angepasst und entsprechend konzeptualisiert werden (Hulsebosch, 2013). Die Effekte von *Gamification* variieren hierbei in Abhängigkeit vom Nutzertypus (Hamari, 2013). Daher ist die Anwendbarkeit in dem vorliegenden HLB-Kontext zu prüfen.

3. Onlinefragebögen zum Einsatz von Web 2.0 in HLB

Um erste Eindrücke zu gewinnen, welche Anforderungen zukünftige Nutzer an eine Expertenplattform mit Web 2.0-Technologien im Kontext von HLB stellen, wurden zwei Onlinefragebögen entwickelt. Diese adressieren jeweils spezifische Zielgruppen: HLB-Experten und Servicetechniker als potentielle Plattformnutzer. Für die HLB-Experten wurden die Einschätzung des Potentials und der Anforderungen an Web 2.0-Technologien im Kontext von HLB erhoben. Bei den Servicetechnikern wurde mittels konkreter Szenarien erfasst, welche Web 2.0-Technologien unter welchen Umständen zum Einholen von Informationen sinnvoll erscheinen. Beide Fragebögen bestanden aus folgenden drei Teilen:

- Demografischen Daten, wie z.B. Erfahrung mit bestimmten Web 2.0-Tools
- Fragen zum erwarteten Nutzenpotential verschiedener Web 2.0-Technologien, wie z.B. Blogs, soziale Netzwerke, Wikis, in HLB (HLB-Experten) sowie als Hilfe bzw. Informationsquelle bei verschiedenartigen technischen Problemen (Servicetechniker)
- Fragen zur Motivationswirkung verschiedener Faktoren, z.B. aus dem Bereich *Gamification*, sich aktiv an solch einer Expertenplattform zu beteiligen

¹ Badges: Hier definiert als „virtuelle Abzeichen“.

² Leaderboards: Hier definiert als „Liste der Teilnehmer mit den höchsten Leistungen / Punkten“.

³ Quests: Hier definiert als „(freiwillig gewählte) Aufgabe / Unternehmung“.

4. Ergebnisse der Onlinebefragungen

Die Onlinebefragung fand vom 30. April bis 17. Juni 2013 statt. An der Befragung der HLB-Experten nahmen 20 wissenschaftliche Mitarbeiter des SFB/TR29 teil (16 männlich, 4 weiblich). Die Altersspanne bewegte sich bei 13 Teilnehmern zwischen 21-30 Jahren, bei 6 zwischen 31-40 Jahren, wobei ein Eintrag ohne Angabe blieb. Als Fachrichtungen waren 6-mal Wirtschafts- und 9-mal Ingenieurwissenschaften sowie 4-mal Informatik vertreten.

An der Servicetechnikerbefragung nahmen 15 Personen teil (alle männlich). Die Altersspanne reichte von 21 bis über 50 Jahren (21-30 Jahre: 5, 31-40 Jahre: 5, 41-50 Jahre: 2, >50 Jahre: 3). Folgende Berufsgruppen waren unter den Servicetechnikern vertreten: 1 Kfz-Mechaniker, 1 Sachverständiger, 1 Werkzeugbauer, 1 Qualitätsplaner sowie 2 Kfz-Meister und 5 Prüfengeure.

Beide Gruppen – HLB-Experten und Servicetechniker – empfinden den Umgang mit neuen Technologien reizvoll und fühlen sich mit dem Internet vertraut. Bei verschiedenen Web 2.0-Technologien nutzen beide Gruppen am häufigsten, d.h. täglich, intelligente Suchsysteme und soziale Netzwerke. Ebenfalls häufig verwenden Servicetechniker Diskussionsforen und HLB-Experten Videoportale.

Im zweiten Teil des Fragebogens folgte die Erhebung des erwarteten Nutzenpotentials verschiedener Web 2.0-Technologien. Hierbei wurden beide Gruppen befragt, wie sinnvoll sie diese im HLB-Kontext (HLB-Experten) sowie als Hilfe bei verschiedenartigen technischen Problemen (Servicetechniker) finden. Es ist festzustellen, dass beide Gruppen die intelligente Suche und Wikis als sinnvoll erachten. Unterschiede zeigen sich darin, dass HLB-Experten zusätzlich Videoportale, die Servicetechniker jedoch Diskussionsforen für sinnvoll erachten. Als generelles Hindernis für die Nutzung werten beide Gruppen, den Zeitaufwand bei der Suche seltener Probleme (Servicetechniker) sowie jenen Zeitaufwand bei der aktiven Beteiligung (HLB-Experten) sowie eine eventuell mangelhafte Aktualität der Informationen. Positiv beurteilt werden die einfache Informationssuche, der Zugang zu Hintergrundinformationen und die Zuverlässigkeit der Informationen.

Im dritten Teil des Fragebogens wurde die Motivationswirkung verschiedener Faktoren, z.B. aus dem Bereich *Gamification*, erfragt. Hierbei gaben die HLB-Experten an, dass die vermutlich höchste Motivationswirkung für eine aktive Beteiligung von Servicetechnikern auf einer Expertenplattform sich aus der Bedingung „Punkte kombiniert mit materiellen Vergünstigungen“ ergibt. Als besonders motivierend wird der Effekt von materieller Belohnung, wie z.B. Rabatte oder einer verlängerten Garantie, eingeschätzt. Demgegenüber werden altruistische Werte, wie z.B. das Gefühl bzw. Wissen anderen Servicetechnikern im HLB durch die eigenen Beiträge zu helfen, als wenig motivierend beurteilt.

Im Kontrast hierzu geben die Servicetechniker an, sich am stärksten durch altruistische Werte zur Teilnahme an der Expertenplattform motiviert zu fühlen. Das bedeutet, dass die Aussicht Kollegen (bekannt und entfernt) mit eigenen Beiträgen

bei technischen Problemen helfen zu können, am stärksten motiviert, z.B. gemachte Produkterfahrungen, mitzuteilen. Ebenfalls ist es den befragten Servicetechnikern wichtig, zur kontinuierlichen Verbesserung des technischen Produkts beizutragen. Untergeordneten Einfluss auf die Nutzung der Expertenplattform haben dagegen Aspekte, wie z.B. die Angst vor mangelnder Datensicherheit, die Gefahr der Konkurrenz mit eigenen Beiträgen zu helfen sowie die Aussicht auf (materielle) Vergünstigungen.

5. Zusammenfassung und Ausblick

In diesem Werkstattbeitrag wurden erste qualitative Ergebnisse zweier Onlinebefragungen zu den Anforderungen an eine Expertenplattform in HLB vorgestellt. Die zu entwickelnde Expertenplattform soll es manuell arbeitenden Servicetechnikern ermöglichen, Expertenwissen über Probleme und Lösungen abzurufen, selbst einzustellen und zu bewerten. Die Anforderungen wurden getrennt für HLB-Experten und Servicetechniker erhoben, um die Besonderheiten im HLB und jene in technischen Kontexten, z.B. bei Wartungsarbeiten, herauszuarbeiten.

Die Ergebnisse zeigen, dass die prototypische Realisierung einer *intelligenten Suche* sowie eines *Wikis* für HLB-Experten und Servicetechniker am sinnvollsten erscheint. Die HLB-Experten erachten zusätzlich die Realisierung eines *Videoportals* und die Servicetechniker ein *Diskussionsforum* als sinnvoll. Als weniger sinnvoll zur Unterstützung technischer Arbeiten im HLB beurteilen beide Gruppen soziale Netzwerke und Blogs.

Hinsichtlich der Motivationswirkung verschiedener Aspekte zeigen sich Unterschiede zwischen den Gruppen. Die HLB-Experten heben die Wichtigkeit von materiellen Vergünstigungen hervor und bezweifeln den Effekt altruistischer Werte auf die Bereitschaft zur aktiven Beteiligung. Die Servicetechniker hingegen betonen die Bedeutung altruistischer Werte, wie z.B. das Wissen, bekannten und entfernten Kollegen mit dem eigenen Beitrag bei einem Problem zu helfen.

Im Projekt folgt nun die prototypische Realisierung der Expertenplattform für HLB mit intelligentem Suchsystem, Wiki, Videoportal und Diskussionsforum.

Literatur

- Anderson, J.R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89, 4, 369-406. DOI: 10.1037//0033-295X.89.4.369.
- Chirumalla, K., Bertoni, M. and Johansson, C. (2013). Experience Feedback Using Social Media: From the Product Lifecycle Phases to the Design Practices. In H. Meier, (Ed.), *Product-Service Integration for Sustainable Solutions*, LNPE, Springer-Verlag, Berlin Heidelberg, 459-471. DOI=10.1007/978-3-642-30820-8_39.
- Deterding, S., Dixon, D., Khaled, R. and Nacke, L. (2011). From game design elements to gamefulness: defining "gamification". In *Proceedings of the 15th*

- international academic mindtrek conference: envisioning future media environments*, mindtrek 2011, ACM New York, NY, 9-15. DOI=<http://dx.doi.org/10.1145/2181037.2181040>
- Ford, D. P. and Staples, D.S. (2005). Perceived Value of Knowledge: Shall I Give You My Gem, My Coal? In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences* (Big Island, Hawaii, January 03 – 06, 2005). 247a. DOI=<http://dx.doi.org/10.1109/HICSS.2005.488>
- Green, M. (2011). Better, smarter, faster: Web 3.0 and the future of learning. *T+D*, 25, 1, 70-72.
- Hamari, J. (2013). Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic Commerce Research and Applications*, Available online 15 February 2013, DOI=<http://dx.doi.org/10.1016/j.elerap.2013.01.004>
- Hulsebosch, M. (2013). Targeting Gamification Applications to Increase User Participation. In *18th Twente Student Conference on IT* (Enschede, The Netherlands, January 25th, 2013).
- Li, R. Y. M. and Poon, S. W. (2011). Using Web 2.0 to Share Knowledge of Construction Safety: The Fable of Economic Animals. *Economic Affairs*, 31, 1 73-79. DOI=<http://dx.doi.org/10.1111/j.1468-0270.2010.02053.x> .
- Luoma, S. and Okkonen, J. (2009). Capturing Competence – Using Wiki for Transferring Tacit Knowledge. In *Proceedings of the European Conference on Intellectual Capital*. 329-336.
- Meier, H., Uhlmann, E., Kortmann, D. (2005). Hybride Leistungsbündel – Nutzenorientiertes Produktverständnis durch interferierende Sach- und Dienstleistungen. *Wt Werkstattstechnik online*, 95 (7), 528-532.
- Mueller, J., Hutter, K., Fueller, J. and Matzler, K. (2011). Virtual worlds as knowledge management platform - a practice-perspective. *Information Systems Journal*, 21, 6, 479–501. DOI=10.1111/j.1365-2575.2010.00366.
- Schmuntzsch, U., Sturm, C., Reichmuth, R. and Rötting, M. (2012). Virtual Agent Assistance for Maintenance Tasks in IPS² - First Results of a Study. In *Advances in Ergonomics in Manufacturing*, Boca Raton, FL: CRC Press. 221-231.
- Tudorache, T., Falconer, S., Noy, N. F., Nyulas, C., Üstün, T. B., Storey, M.-A. and Musen, M. A. (2010). Ontology Development for the Masses: Creating ICD-11 in WebProtégé. In *Knowledge Engineering and Management by the Masses*, 6317, 74-89. DOI=10.1007/978-3-642-16438-5_6.
- Uhlmann, E., Meier, H., Stelzer, C., Geisert, C., Bochnig, H. and Sadek K. (2008). Design of PSS Based on Customer Requirements. *First International Seminar on PSS*, Shaker, Bochum.

Acknowledgements

We thank the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) for funding this research within the Transregional Collaborative Research Project TRR 29 on Industrial Product-Service Systems – dynamic interdependencies of products and services in production area.

Cross-Cultural Usability: Eine Untersuchung wie Usability von deutschen vs. russischen Nutzern verstanden wird

Eva-Maria Skottke¹, Jana Warkentin und Tatiana Krott²

Schlüsselwörter: Cross-Cultural Usability, HCI, Ergonomie

Zusammenfassung

Häufig wird ein technisches System in einen anderen Kulturkreis exportiert als in welchem es entwickelt wurde. Durch unterschiedliche kulturelle Orientierungssysteme kann die Folge dann sein, dass das Modell des Entwicklers hinsichtlich Nutzungsanforderungen nicht optimal zu den Benutzern der Zielkultur passt. Die im Entwicklungsland erreichten Usability-Anforderungen können somit in der Zielkultur eine völlig andere Bewertung erreichen. Interkulturelles Usability Engineering sollte daher vorhandene kulturelle Modelle erheben und Kulturanforderungen schon bei der Produktgestaltung berücksichtigen. Die hier vorgestellte Studie zeigt auf, welche unterschiedlichen Nutzeranforderungen zwischen deutschen und russischen Usern hinsichtlich Textverarbeitungs- und E-Mail-Programmen bestehen.

Ausgangsbefunde zur Cross-Cultural Usability

Der Kulturbegriff ist sehr vieldeutig. In den zahlreichen Definitionen findet man jedoch zumeist diese drei gemeinsamen Aspekte (Triandis, 2010): Kultur entsteht aus der Wechselwirkung zwischen den Menschen und ihrer Umgebung, Menschen, die einer gemeinsamen Kultur zuzurechnen sind, besitzen Gemeinsamkeiten, die sie von anderen Kulturen abgrenzen und Kultur wird von Menschen an andere Menschen weitergegeben. Jeder Mensch besitzt also eine kulturelle Identität, geprägt durch geografische, botanische Bedingungen (Behringer, 2009) oder auch genetische Veranlagungen der Menschen, die einen Einfluss auf die Kommunikationsprozesse haben, beispielsweise bei der Mensch-Computer-Interaktion. Nahelegend ist somit die große Bedeutung der Berücksichtigung von kulturellen Unterschieden in benutzerfreundlich gestalteten Computersystemen. Dieselbe Information in derselben Darbietung kann nämlich von Angehörigen unterschiedlicher Kulturen unterschiedlich aufgenommen und verstanden werden, was zurückzuführen ist auf die Entwicklung einer jeden Kultur von eigenen Werten und Symbolen, die ganz unterschiedliche Bedeutungen und Interpretationen haben können. Missverständnisse können zu fehlerhaften Handlungen führen, was sowohl Folgen für die Produktivität des Nutzers, als auch für seine Akzeptanz (Frustration) haben kann.

¹ Hochschule für Medien, Kommunikation und Wirtschaft (HMKW), Campus Köln

² RWTH Aachen University

Möglicherweise wird in Zukunft eine der Voraussetzungen für eine benutzerorientierte Software sein, dass diese auch kulturorientiert ist.

In mehreren Usability-bezogenen Studien wurden bereits kulturelle Unterschiede untersucht und bestätigt. Besonders intensiv wurden kulturelle Nutzerbesonderheiten in asiatischen Ländern wie China oder Japan und in Deutschland und den USA erforscht. Hier sei beispielhaft eine Studie von Duda, Schießl und Nüsperling (2008) angeführt, die das freie und aufgabenbezogene Surfen auf Webseiten durch deutsche und japanische Nutzer verglichen. Mittels Blickbewegungsaufzeichnung und der subjektiven Bewertung der Webseiten konnten deutliche Unterschiede zwischen den Kulturen festgestellt werden. Die japanischen Probanden waren aufgrund ihres komplizierten Schreibsystems und ihrer eher visuell orientierten Kultur eine hohe Komplexität gewöhnt und bewerteten Websites mit sehr viel Inhalt besser als die deutschen Probanden. Die deutschen Nutzer bevorzugten eher einfach strukturierte Webseiten mit einer klaren Positionierung der wichtigsten Inhalte.

Eine Analyse zu den Nutzerpräferenzen beim Gebrauch von Smartphones zwischen amerikanischen und koreanischen Probanden (Shin, 2011) zeigte, dass ein unterschiedliches Wertesystem hinsichtlich der Präferenz und Interaktion mit Smartphones in Bezug auf Usability und Ästhetik vorherrschte. Die koreanischen Probanden wurden bei ihrer Beurteilung stärker durch ästhetische Aspekte beeinflusst, wo hingegen die Nutzer aus den USA die gegenteilige Präferenz zeigten und die funktionalen Aspekte höher bewerteten. Die Nutzungsgewohnheiten waren allerdings auch sehr unterschiedlich: Während die Koreaner sehr häufig das Smartphone für so genannte "hedonic services" wie Musik, Videos und Spiele nutzten, fand man bei den Amerikanern eher die Nutzung von funktionalen oder informativen Diensten, was die Ergebnisse sicherlich beeinflusst hat.

Fragestellung und Herleitung der Hypothesen

Ausgangspunkt der eigenen Arbeit bildet eine Studie von Frandsen-Thorlacius und Kollegen (2009), die untersuchten, wie Usability von chinesischen und dänischen Nutzern verstanden wird und die von unserer Arbeitsgruppe mit deutschen und russischen Probanden ergänzt wurde. In der Ausgangsstudie wurde 412 Nutzern aus China und Dänemark ein Fragebogen vorgelegt, in dem die Wichtigkeit der unterschiedlichen Aspekte von Usability von Textverarbeitungs- und E-Mail-Programmen bewerten sollten. Außerdem sollten sie in einem Vergleichsrating den wichtigeren Aspekt aus zwei genannten auswählen. Die abgefragten Aspekte waren Effektivität, soziale Aktivität, Effizienz, Zufriedenheit, Einfachheit der Benutzung, Spaß, Abwesenheit von Frustration bei der Benutzung und das Erscheinungsbild. Die Analyse ergab, dass für die chinesischen Nutzer das Erscheinungsbild, Zufriedenheit und Spaß wichtiger als für die dänischen Nutzer war. Die Dänen interessieren sich mehr für die Effektivität, Effizienz und die Abwesenheit von Frustration bei der Benutzung. Außerdem zeigten die Ergebnisse, dass auch der Typ des Programms eine Rolle spielt. Für dänische Teilnehmer ist Effektivität sig-

nifiktant wichtiger bei einem Textediting- als bei einem E-Mail-Programm. Die Ergebnisse der Studie sind auszugsweise in Abbildung 1 dargestellt.

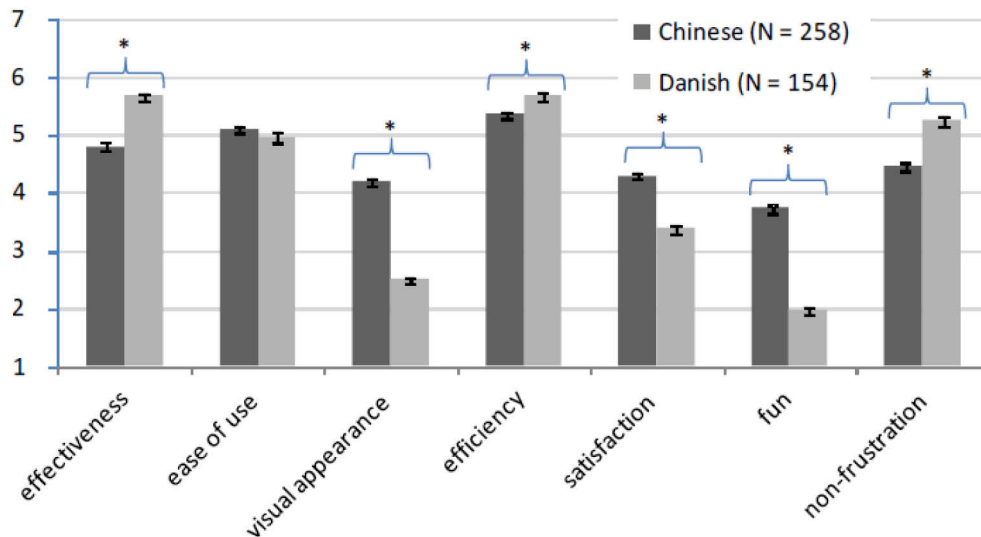


Abb. 1: Auszughafte Ergebnisse der Studie von Frandsen-Thorlacius et al. (2009): Mittlerer Antwortwert bei den sieben Skalen

Ausgehend von diesen Befunden stellte sich die Frage, wie ein Vergleich zwischen deutschen und russischen Nutzern ausfallen würde, da die zunehmende Globalisierung bzw. der Export der Warenkategorie "Datenverarbeitungsträger" aus Deutschland in die Russische Föderation stark zunimmt. Es wurden bis jetzt keine Untersuchungen mit russischen Stichproben durchgeführt, die deren Bedürfnisse hinsichtlich Usability erforscht haben (vgl. z.B. Sergeev, 2008).

In der kulturvergleichenden Persönlichkeitsforschung allerdings findet man interessante Forschungsergebnisse, die hier zur Hypothesengenerierung aufgegriffen wurden. Eine bekannte kulturelle Untersuchung stammt von Hofstede & Hofstede (2005): Den Autoren zufolge ist in Deutschland das Ausmaß an Unsicherheitsvermeidung recht hoch - neue Produkte und Technologien werden erst nach einer Zeit akzeptiert. Deutschland gehört laut der Autoren außerdem zu den Ländern mit niedriger kontextabhängigkeit. Neben verbalen Informationen gibt es nur wenig andere Aspekte, die eine entscheidende Rolle spielen (im Gegensatz zu z.B. in China, wo Aspekte wie Höflichkeit, Mimik, Gestik usw. von hoher Bedeutung sind). Übertragen auf die Mensch-Computer-Interaktion bedeutet dies, dass die deutschen Nutzer eher knappe und kurze Meldungen bevorzugen, auch wenn es auf Kosten der Höflichkeit geht. Außerdem bevorzugen Deutsche ein schlichtes Design, mit weniger Abbildungen und Farben (vgl. auch Boreum, Inseong & Jinwoo, 2006). Russland weist in der Hofstede-Studie eine relativ hoch ausgeprägte „Machtdistanz“ auf. Hier bestehen im Gegensatz zu Deutschland stark ausgeprägte Hierarchien. Der Vorgesetzte dominiert und wird widerstandslos akzeptiert. Personen an machthabenden Positionen genießen mehr Privilegien und haben ei-

nen eher patriarchalischen Führungsstil. Weniger Macht wird gleichgesetzt mit weniger Rechten. Unsicherheitsvermeidung ist auch in Russland sehr hoch ausgeprägt. Diese Dimension wirkt sich auf Angst vor neuartigen Situationen und Adaptationsschwierigkeiten aus. Die hierarchische Orientierung und die damit zusammenhängende hohe Machtdistanz begründen Trommsdorff & Schuchardt (1998) als Folge einer wechselwirkenden Verstärkung von Jahrhunderte lang praktizierten christlich-orthodoxen Lehren und der im sowjetischen Staat praktizierten monohierarchischen Bürokratie.

Eine weitere Methode zur Erkennung der kulturellen Unterschiede stellt die Analyse von nationalen Fremd- und Selbststereotypen dar (siehe Rösch, 2002). Die Analyse liefert ein rein subjektiv wahrgenommenes Abbild der Kultur.

Rösch (2002) führte eine Befragung mit 186 russischen und 97 deutschen Studenten zu ihren Auto- und Heterostereotypen durch. Ein Autostereotyp ist ein Urteil, welches die befragte Personengruppe über sich selbst bildet und fällt. Ein Heterostereotyp ist ein Urteil über eine andere Personengruppe, die es zu bewerten gilt. In der Befragung von Rösch wurde nach freien Assoziationen gefragt, es wurde also keine vordefinierte Eigenschafts- und Aussagenliste vorgelegt. Die russische Gruppe bezeichnete sich in dieser Studie beispielsweise als gutmütig, fröhlich, faul, hilfsbereit, arbeitsam, einfach und trinkfreudig bezeichnet, während die Fremdwahrnehmung durch die Deutschen Eigenschaften wie: trinkfreudig, gastfreundlich, traditionsbewusst, fröhlich, feiern gern, bescheiden, familiär beinhalten. Die deutsche Gruppe bezeichnete sich selbst beispielsweise als pünktlich, ordentlich, arrogant/ überheblich, strebsam/ zielstrebig, bürokratisch und fleißig, während sie von den Russen als klug, pünktlich, arbeitsam, korrekt, sparsam/wirtschaftlich, gastfreundlich gesehen wurde. Aufgrund dieser Recherche wird hypothetisiert, daß „Effektivität“ und „Effizienz“ für deutsche Benutzer wichtiger ist als für russische Benutzer. Für russische Benutzer ist die „Erscheinung“ und „Spas“ wichtiger als für deutsche Benutzer.

Methode

An der Fragebogenstudie nahmen 137 Probanden (65 Deutsche, 72 Russen) teil. Die Probanden waren zwischen 18 und 53 Jahre alt ($M=26$), das Geschlecht war bei beiden Kulturen ausbalanciert. Fast 90% der Befragten gaben an, ihr E-Mail-Programm täglich zu benutzen, 61% nutzen ihr Textverarbeitungsprogramm täglich, 34% mindestens wöchentlich. Jeder Proband lebte ausschließlich in seinem Land und hatte noch keinen Kontakt mit E-Mail- oder Textverarbeitungsprogrammen des vergleichenden Landes.

Der Fragebogen zur Bewertung der Wichtigkeit verschiedener Usabilityaspekte (z.B. Aspekte der ISO 9241-11: Effektivität, Effizienz, Zufriedenheit; sowie Spas, Non-Frustration, u.a.) wurde aus der Studie von Frandsen-Thorlacius (s.o.) übernommen und ins Deutsche und Russische übersetzt. Er umfasste absolute Fragen zu Aspekten der Usability, wobei jeder Aspekt durch eine Skala mit 4 Items ge-

messen wurde. Jedes Item wurde mit einer 7-stufigen Skala versehen von 1: "nicht wichtig" bis 7: "extrem wichtig". Zusätzlich gab es noch vergleichende Fragen, deren Ergebnisse in diesem Artikel jedoch nicht aufgegriffen werden.

Die Inter-Item-Reliabilität der Fragen wurde für jede Skala mittels Cronbach's Alpha getestet. Eine Skala wird üblicherweise als reliabel angesehen, wenn der Alpha-Wert die Schwelle von 0.7 überschreitet. Alle absoluten Skalen erreichten ein Alpha-Level höher als 0.7.

Ergebnisse

Die Ergebnisse der Befragung zeigen, dass bei E-Mail-Programmen für die Deutschen und Russen wichtige Faktoren Effektivität, Einfachheit der Benutzung und Effizienz sind. Signifikante Unterschiede zwischen den Kulturen weisen hier die Faktoren Effektivität (Deutsche: 5,3; Russen: 4,7), Erscheinung (Deutsche: 2,6; Russen: 3,4), Effizienz (Deutsche: 6,1; Russen: 5,5), Zufriedenheit (Deutsche: 3,6; Russen: 4,3), Spass (Deutsche: 2,15; Russen: 2,8) sowie "Non-Frustration" (Deutsche: 5,5; Russen: 4,4) auf (vgl. Abb. 2).

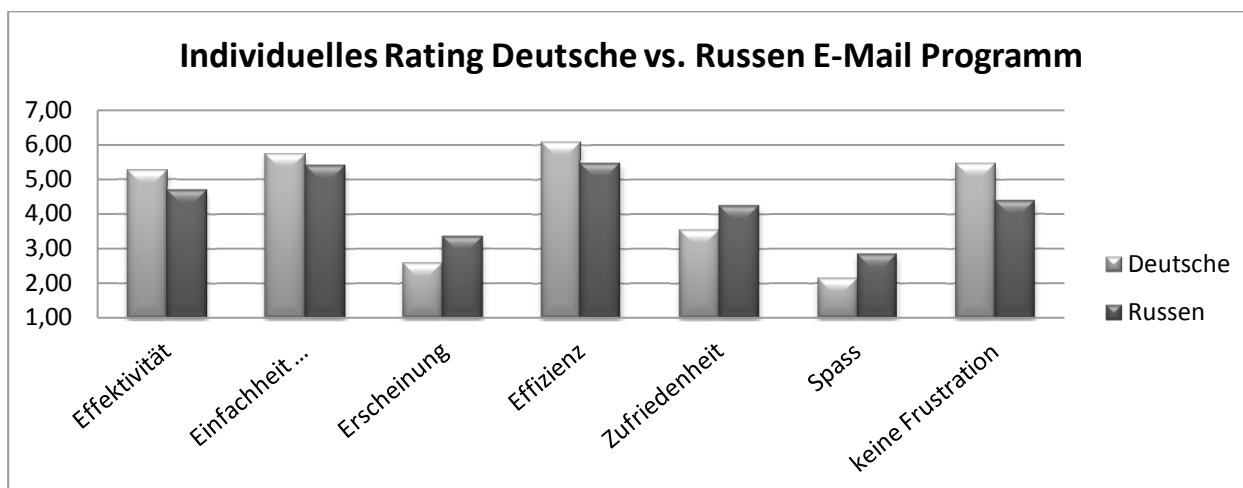


Abb. 2: Vergleich des individuellen Ratings zwischen Deutschen und Russen bei E-Mail Programmen: Mittlerer Antwortwert bei den sieben Skalen

Für das Wichtigkeitsrating zu Textediting-Programmen ergab sich eine hohe Priorität für beide Kulturen für die Kategorien Einfachheit der Benutzung, Erscheinungsbild sowie Effizienz. Unterschiede zwischen den Kulturen konnten in den Kategorien Effektivität (Deutsche: 6,0; Russen: 5,2), Effizienz (Deutsche: 6,05; Russen: 5,3) und "Non-Frustration" (Deutsche: 5,6; Russen: 4,2) gefunden werden (vgl. Abb. 3). Insgesamt zeigen die Daten, dass die Hypothese, dass Effektivität und Effizienz für deutsche Nutzer wichtiger sind als für russische Nutzer, bestätigt werden konnte - dies sowohl für E-Mail- als auch für Textediting-Programme. Die Annahme, dass die äußere Erscheinung des Programms und der

Spaßfaktor bei den russischen Nutzern signifikant höher sind, konnte zumindest für die Bewertung von E-Mail-Programmen nachgewiesen werden.

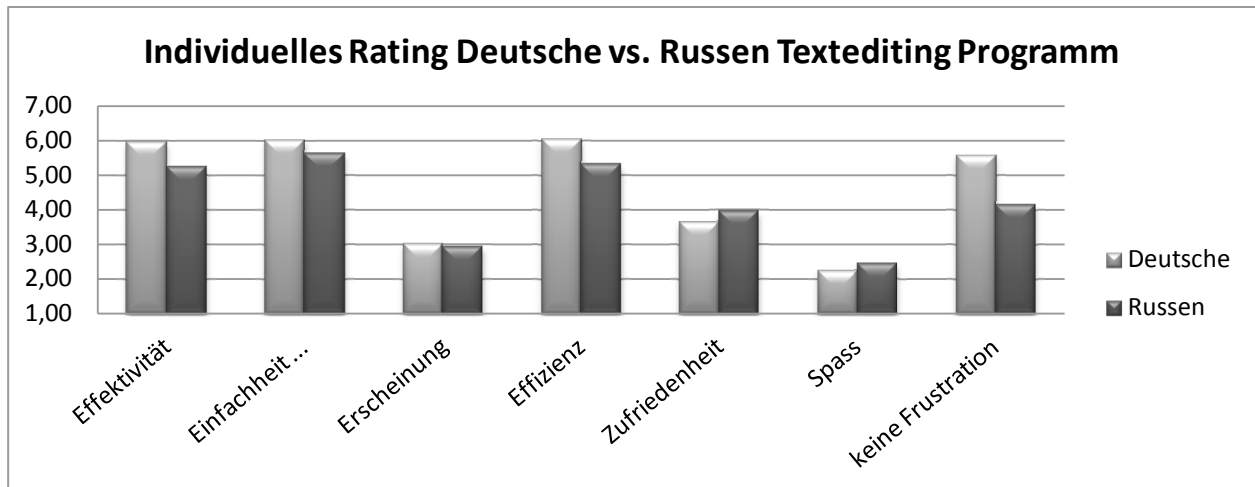


Abb. 3: Vergleich des individuellen Ratings zwischen Deutschen und Russen bei Textediting-Programmen: Mittlerer Antwortwert bei den sieben Skalen

Zusammenfassung und Handlungsempfehlungen

Die Fragebogenstudie mit deutschen und russischen Programmnutzern konnte zeigen, dass in den beiden Kulturen verschiedenen Usabilityaspekten bei der Nutzung von E-Mail- und Textediting-Programmen signifikant unterschiedliche Gewichtung beigemessen wird. Beide Nutzergruppen gewichten Effektivität, Effizienz und Einfachheit bei der Benutzung stark, die ersten beiden genannten Aspekte liegen bei den deutschen Nutzern jedoch noch deutlich höher in der Gewichtung. Die deutschen Befunde liegen damit sehr nah an denen der Dänen, die russischen sehr nah an denen der Chinesen (vgl. Ausgangsstudie von Frandsen-Thorlacius et al., 2009).

Für die Gestaltung von Software und Webseiten können diese Ergebnisse zu mehr individuellen und kulturangepassten Produkten führen, welches sich möglicherweise in den Absatzzahlen und der Zufriedenheit der User bemerkbar macht.

Literatur

- Behringer, W. (2009). Kulturgeschichte des Klimas: Von der Eiszeit zur globalen Erwärmung. München: C. H. Beck.
- Boreum, C., Inseong L., & Jinwoo K. (2006). Culturability in mobile data services: A qualitative study of the relationship between cultural characteristics and user-experience attributes. *International Journal of Human-Computer Interaction*, 20(3), 171–203.
- Duda, S., Schießl, M. & Nüsperling, S. (2008). See the World with different Eyes. Comparison of Website Usage in Germany and Japan. In: *planung & analyse Market Research Special Engl. Edition 2008: Asia and Middle East*. S. 14-18.

- Frandsen-Thorlacius O., Hornbæk K., Hertzum M., & Clemmensen T. (2009). Non-universal usability? A survey of how usability is understood by Chinese and Danish users. *Proceedings of the CHI 2009*. New York: ACM.
- Hofstede, G., & Hofstede G. J. (2005). *Cultures and organizations: software of the mind*. New York: McGraw-Hill.
- Rösch, O. (2002). Mit Stereotypen leben? Wie Deutsche und Russen sich heute sehen. In O. Rösch (Eds.) *Interkulturelle Kommunikation in Geschäftsbeziehungen zwischen Russen und Deutschen. Beiträge aus Wissenschaft und Praxis* zum 2. Wildauer Workshop „Interkulturelle Kommunikation“ (pp. 53-64). Berlin: News &Media.
- Sergeev S.F. (2008). *Инженерная психология и эргономика*. Москва: НИИ школьных технологий.
- Shin, D. (2012). Cross-analysis of usability and aesthetic in smart devices: what influences users' preferences? *Cross Cultural Management*, 19 (4); 563-587.
- Triandis, H.C. (2010). Culture and Psychology: A History of the Study of Their Relationship. In S. Kitayama & D. Cohen (Eds.) *Handbook of Cultural Psychology* (pp. 59-76). New York: The Guilford Press.
- Trommsdorff, V. & Schuchardt, C. (1998). *Transformation osteuropäischer Unternehmen: Grundlagen – Rahmenbedingungen – Strategien*. Wiesbaden: Gabler.

Eignung unterschiedlicher Versuchsumgebungen für die Erfassung der wahrgenommenen Qualität von Flugzeugkabinen

Lennart Weber¹ und Inge Wetzel¹

Schlüsselwörter: Perceived Quality, Versuchsumgebungen, Virtuelle Realität

Zusammenfassung

Wahrgenommene Qualität (Perceived Quality, PQ) ist das Gesamturteil eines Kunden über ein Produkt, das anhand von Bewertungen einzelner Aspekte gebildet wird. Die PQ von Flugzeugkabinen zu untersuchen ist ein Forschungsbereich bei EADS Innovation Works. Da das Urteil der Passagiere über die Kabine durch sensorische Eindrücke erfolgt, wird ein psychophysikalisch orientierter, experimenteller Forschungsansatz verfolgt. Durch systematische Variation physikalischer Parameter der Kabine wird deren Auswirkung auf die subjektive Wahrnehmung untersucht. Die physikalisch korrekte Darstellung der Umwelt ist dabei eine große Herausforderung, da sie den eigentlichen Untersuchungsgegenstand bildet und keine Darstellungsumgebung für eine bestimmte Aufgabe oder ein Prinzip. Es werden unterschiedliche Versuchsumgebungen verwendet: Labor mit Computer, physisches Mock-Up und Virtuelle Realität (Powerwall, Head-Mounted-Display, CAVE), welche nach drei Kriterien bewertet werden können: Kontrollierbarkeit, Veränderbarkeit und Realitätsnähe. Die vorgestellten Arten der Versuchsumgebung weisen diesbezüglich große Vor- und Nachteile auf, keine wird jedoch allen Kriterien in ausreichendem Maße gerecht.

Labor mit Computer

In einer einfachen Laborumgebung können unterschiedliche Entwürfe von Flugzeugkabinen präsentiert und Blickbewegungen mit einem Remote-Eye-Tracker aufgezeichnet werden. Das digitale Labor zeichnet sich durch eine sehr hohe Kontrollierbarkeit aus, da Gestaltungsmöglichkeiten der Umgebung schnell umgesetzt werden können. Die Veränderbarkeit ist jedoch stark eingeschränkt: Die Stimuli (Kabinenlayouts) können nur mit Bildbearbeitungssoftware verändert werden. Realitätsnähe ist kaum gegeben: Versuchspersonen tun sich schwer in der Vorstellung der entsprechenden realen Umgebung. Es bereitet ihnen sichtlich große Mühe, aus ihrer Sicht korrekte Antworten auf die Als-Ob-Fragen zu geben. Die Bereitstellung eines realitätsnahen Kontextes (Erhalt eines Tickets, Platzsuche, Verstauen des Gepäcks) gestaltet sich als äußerst schwierig bis gar nicht durchführbar. Eine weitere bedeutende Einschränkung besteht in der Beschränkung auf die visuelle und die akustische Domäne. Selbst in der visuellen Domäne ist nur ein sehr einge-

¹ EADS Innovation Works Hamburg

schränktes Wahrnehmungsfeld vorhanden, da das Stimulusmaterial durch die Abmessungen des Monitors beschränkt ist.

Eine Laborumgebung kann andererseits hervorragend als Pre-Test-Umgebung genutzt werden. Mit sehr wenig Aufwand können grafische Prototypen erstellt und getestet werden. Die Auswertung eines solchen Pre-Test erlaubt es, erste Urteilstendenzen vorwegzugreifen.

Physisches Mock-Up

In einem Kabinen-Mock-Up kann die Kabinenumgebung mit all ihren sensorischen Eigenschaften (Geräusche, Materialeigenschaften etc.) auf realistische Weise dargestellt werden. So lassen sich Veränderungen dieser physikalischen Aspekte exakt messen und ihre Auswirkung auf PQ untersuchen. Mock-Ups zeichnen sich durch einen hohen Grad an Realitätsnähe aus, da dort reale Kabinenteile eingesetzt werden. Ganze Sektionen – auch als "Tonnen" bezeichnet – können direkt nachgebaut werden (siehe Abbildung 1).



Abbildung 3: Airbus-Mock-Up

Außerdem werden alle sensorischen Modalitäten gleichermaßen bedient. In einem Mock-Up haben Versuchspersonen die größtmögliche Bewegungsfreiheit, und die aufgabengebundene Interaktion mit der Kabine kann ohne Einschränkungen erfolgen: Es ist beispielsweise möglich, ein Einstiegsszenario inklusive Gepäck-Verstauen (haptische Interaktion) umzusetzen.

Große Nachteile bieten die geringe Kontrollierbarkeit und die geringe Veränderbarkeit. Dies bedeutet eine starke Einschränkung in der Wahl der zu manipulierenden physikalischen Eigenschaften und der Anzahl der realisierbaren Ausprägungsstufen. Bestimmte Aspekte der Umgebung sind gar nicht veränderbar, da nur eine einzige Variante existiert. Weiterhin ist der Bau vollkommen neuer Kabinenteile mit sehr hohen Kosten verbunden.

Virtuelle Realität

Experimente werden in unterschiedlichen Umgebungen aus dem Bereich der Virtuellen Realität (VR) durchgeführt: Powerwall, Head-Mounted-Display und Cave Automatic Virtual Environment (CAVE).

Eine Powerwall besteht aus einer Rückprojektionswand, die von außen durch einen oder mehrere Projektoren angestrahlt wird. Die visuelle Darstellung kann dabei sowohl monoskopisch als auch stereoskopisch über eine Polarisationsbrille erfolgen.

Ein Head-Mounted-Display ist eine Vorrichtung, die am Kopf der Versuchsperson befestigt wird. Die visuelle Darstellung erfolgt über zwei Displays, die sich direkt vor den Augen befinden. Für ein Auge ist jeweils ein Display vorgesehen. Die Versuchsperson verfügt im Gegensatz zu anderen Technologien über keine Außensicht. Das Gerät ist häufig mit Track-Balls versehen, die es einem kamera-basierten Tracking-System ermöglichen, die Lage des Geräts im Raum zu lokalisieren und die Darstellung in den Displays darauf auszurichten (siehe Abbildung 2).



Abbildung 4: Versuchsperson mit HMD und Tracking-System

Unter einer CAVE wird ein Raum verstanden, der – anders als eine Power Wall – aus mehreren Rückprojektionswänden besteht. Diese Umgebung ermöglicht ein großes Maß an Bewegungsfreiheit. Über eine Polarisationsbrille wird das Stimulusmaterial aufgenommen. Diese ist ebenfalls mit Track Balls ausgestattet, damit die Bilder auf die Position der Brille im Raum ausgerichtet werden können.

Umgebungen in virtueller Realität zeichnen sich durch eine hohe Kontrollierbarkeit und eine besonders große Veränderbarkeit aus: Das Kabinenlayout kann mittels 3D-Visualisierungssoftware genau festgelegt werden. Gleichzeitig besteht eine beinahe grenzenlose Menge an Freiheitsgraden der Manipulation (Geometrie, Oberflächen, Licht). Doch auch VR-Umgebungen können keine exakte Simulation der Wirklichkeit liefern: Physikalische Lichtmessungen zeigen, dass zwischen theoretisch simulierter und tatsächlich realisierter Farbtemperatur z.B. eine Diskre-

panz von ca. 2000 Kelvin existiert. Ein weiteres Problem betrifft die Gleichmäßigkeit der Visualisierung. Bei Verwendung von Projektionstechniken entsteht ein sogenannter Hotspot-Effekt: In der Mitte herrscht eine verhältnismäßig hohe Leuchtdichte, die zur Peripherie stark abnimmt (von über 400 cd/m² bis zu 30 cd/m²; siehe Abbildung 3).

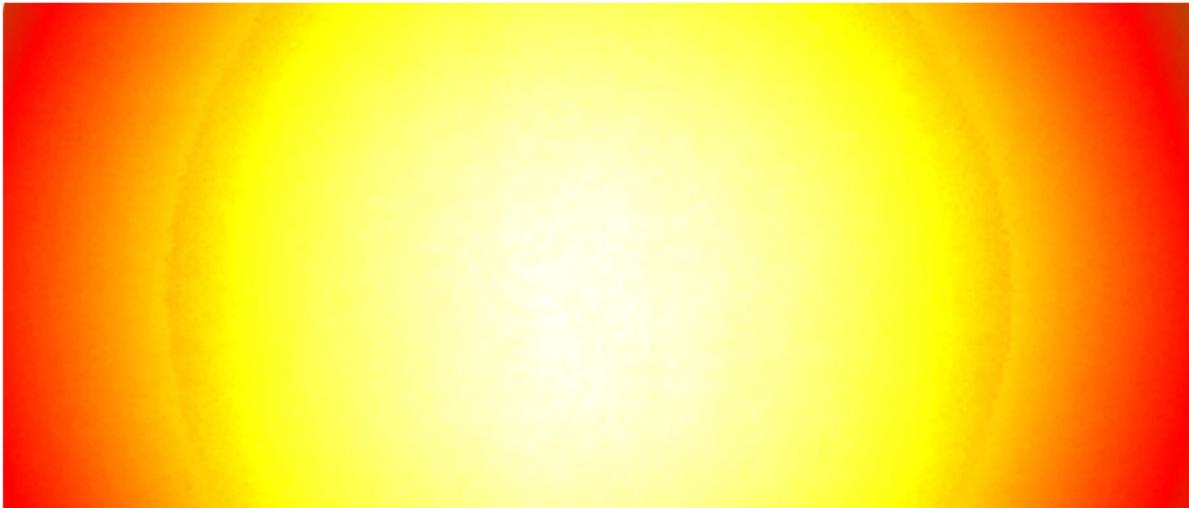


Abbildung 5: Leuchtdichteverteilung eines weißen Bildes an einer Powerwall

Die Integration physischer Elemente in die Szene stellt eine weitere große Herausforderung dar. Dennoch können mittels VR stark immersive Eindrücke erzeugt werden (Armbrüster et al., 2008), und das Antwortverhalten ähnelt stark dem in Konfrontation mit echten Produktinstanzen (Berneburg, 2007, 2008). VR ermöglicht die Anwendung komplexer Versuchsdesigns: Es ist problemlos möglich, vier- oder mehrfaktorielle Versuchspläne mit mehreren Messwiederholungsfaktoren zu implementieren. Eine weitere Möglichkeit der Testung von Varianten, die sich in Bezug auf mehrere Eigenschaften unterscheiden bietet die Methodenklasse der Choice-Based Conjoint Analyse: Diese besteht in einem simulierten Entscheidungsprozess, in dem sich die Versuchspersonen in Sets von zwei oder mehr Alternativen jeweils für eine Variante entscheiden müssen.

Fazit

Jede der aufgeführten Versuchsumgebungen zeichnet sich durch klare Vorteile und Nachteile für die Erforschung der PQ von Flugzeugkabinen aus. Sie sind in Bezug auf die hier angewendeten Kriterien komplementär (siehe Tab. 1).

Tab. 2: Vorzüge und Nachteile der vorgestellten Versuchsumgebungen

Kriterium	Labor	Mock-Up	Virtual Reality
Kontrollierbarkeit	+	-	+
Realitätsnähe	-	+	0
Veränderbarkeit	0	-	+

Es ist dennoch nur schwer möglich, sie einander ergänzend einzusetzen. Bis zur physikalisch korrekten, multimodalen und flexiblen Kabinen-Versuchsumgebung ist es noch ein sehr weiter Weg.

Literatur

- Armbrüster, C., Wolter, M., Kuhlen, T., Spijkers, W., & Fimm, B. (2008). Depth Perception in Virtual Reality: Distance Estimations in Peri- and Extrapersonal Space. *CyberPsychology & Behavior*, 11(1), 9–15.
- Berneburg, A. (2007). Interactive 3D Simulations in Measuring Consumer Preferences: Friend or Foe to Test Results?. *Journal of Interactive Advertising*, 8(1), 1–13.
- Berneburg, A. (2008). Presence in a Three-Dimensional Test Environment: Benefit or Threat to Market Research? *Journal of Virtual Reality and Broadcasting*, 5.

Kontrasteffekte bei der Benutzung von Likelihood Alarmsystemen

Magali Baloud¹, Dietrich Manzey²

Schlüsselwörter: Likelihood Alarmsysteme, Reliabilität, Vertrauen, cry-wolf Effekt

Zusammenfassung

Likelihood Alarmsysteme (LAS) stellen eine Alternative zu Binären Alarmsystemen (BAS) dar. LAS besitzen mindestens drei Stufen: üblicherweise einen „grünen Bereich“ für einen fehlerfreien Zustand, eine „Warnstufe“ und eine „Alarmstufe“. Die letzten beiden geben Informationen darüber, wie sicher die Diagnose des Systems ist. Frühere Studien zeigen, dass Personen eher auf die Alarmstufe eines LAS reagieren als auf die Alarmstufe eines BAS mit gleicher Reliabilität. Der Grund könnte ein Kontrasteffekt durch die Existenz einer Warnstufe mit niedrigerer Reliabilität sein, welche die wahrgenommene Reliabilität der Alarmstufe implizit erhöht. Diese Hypothese wurde in einem Laborexperiment mit $n = 29$ Versuchsteilnehmern untersucht. Die Ergebnisse zeigen, dass die Probanden die Alarmstufe des LAS zwar nicht als reliabler einstufen als die des BAS, ihr jedoch mehr vertrauen. Die Ergebnisse werden am Ende des Papers ausführlicher diskutiert.

Einleitung

In Situationen, in denen mehrere Aufgaben gleichzeitig bearbeitet werden müssen oder in denen aus anderen Gründen eine hohe Arbeitsbelastung vorliegt, können Alarmsysteme äußerst hilfreich sein – so etwa im Bereich der Luftfahrt, im medizinischen Kontext oder in der Prozessindustrie. Dabei fungieren sie als Mediator zwischen einem bestimmten Prozess und dem menschlichen Operateur, indem sie den aktuellen Status des Prozesses überwachen und den Operateur über diesen informieren. Ihr Einsatz soll dazu beitragen, dass kritische Ereignisse nicht übersehen werden. Das am häufigsten eingesetzte Alarmsystem ist das Binäre Alarmsystem (BAS), das den Operateur in einer binären Form informiert: Entweder es meldet ein kritisches Ereignis (rot) oder nicht (grün).

Im Idealfall gibt ein Alarmsystem nur dann einen Alarm aus, wenn tatsächlich ein kritisches Ereignis vorliegt. In der Praxis ist das jedoch nicht immer der Fall. Alarmsysteme tendieren dazu, eine große Zahl von False Alarms zu produzieren, d.h. sie geben einen Alarm aus, obwohl kein kritisches Ereignis vorliegt. Die so erzeugte, große Zahl an False Alarms hat Auswirkungen auf das mentale Modell, das Operateure von der Reliabilität des Alarmsystems aufbauen. Dieses mentale Modell hängt vor allem vom so genannten Positive Predictive Value (PPV) (Getty,

1. Technische Universität Berlin, Zentrum Mensch-Maschine-Systeme, Graduiertenkolleg prometei
2. Technische Universität Berlin, Institut für Psychologie und Arbeitswissenschaften

Swets, Pickett, & Gonthier, 1995) ab. Der PPV ist die bedingte Wahrscheinlichkeit für ein kritisches Ereignis unter der Bedingung, dass ein Alarm vorliegt. Ein PPV von 0.6 bedeutet also, dass es im Falle eines Alarms nur mit einer Wahrscheinlichkeit von 60% tatsächlich ein kritisches Ereignis gibt.

Das Hauptproblem ist nun, dass Operateure umso weniger einem Alarmsystem vertrauen, je niedriger sein PPV ist (Madhavan & Wiegmann, 2006). In Bezug auf das Verhalten der Operateure führt dies zum „cry-wolf effect“ (Breznitz, 1984; Getty et al. 1995; Bliss et al. 1995): Die Personen neigen dann dazu, langsamer auf einen Alarm zu reagieren oder diesen sogar vollständig zu ignorieren. Ein Verlust an Sicherheit und Produktivität ist die Folge (Lee & See, 2004).

Eine mögliche Lösung für dieses Problem ist die Verwendung von Likelihood Alarmsystemen (LAS). Diese Art von Alarmsystem besteht aus drei oder mehr Stufen, wobei jede Stufe mit einer unterschiedlichen Wahrscheinlichkeit verbunden ist, mit der ein kritisches Ereignis vorliegt (Bustamante & Bliss, 2005). Mit anderen Worten besitzt jede Stufe ihren eigenen PPV. Ein LAS mit drei Stufen könnte beispielsweise zusammengesetzt sein aus einer Alarmstufe, bei der mit hoher Wahrscheinlichkeit ein kritisches Ereignis vorliegt (hoher PPV), einer Warnstufe, die eine niedrigere Wahrscheinlichkeit für ein kritisches Ereignis repräsentiert (niedriger PPV), sowie aus einem grünen Bereich, der anzeigen soll, dass der Prozess gerade fehlerfrei abläuft.

Das Ziel des LAS ist es nun, das Verhalten des Operators so zu beeinflussen, dass er häufiger auf Alarme reagiert, die mit einer hohen Wahrscheinlichkeit für ein kritisches Ereignis verbunden sind, und seltener auf Alarme, bei denen die Wahrscheinlichkeit für ein kritisches Ereignis gering ist. Das bedeutet, dass sich der cry-wolf effect zu großen Teilen auf die Warnstufe beschränken sollte, da hier die Wahrscheinlichkeit für ein kritisches Ereignis geringer ist, während Personen die Alarmstufe häufiger befolgen sollten. Daraus resultiert außerdem, dass Operateure auch häufiger auf Hits – also auf Alarme, die mit einem tatsächlichen kritischen Ereignis einhergehen – und seltener auf False Alarms reagieren. Bustamante und Bliss (2005) haben beispielsweise gezeigt, dass der Einsatz eines LAS den cry-wolf effect abschwächen kann. Insgesamt haben die Versuchspersonen seltener auf das Alarmsystem reagiert, gleichzeitig aber häufiger auf Hits geantwortet. Offenbar scheint ein LAS die Genauigkeit der Entscheidungen zu verbessern (Bustamante, 2008; Wiczorek & Manzey, 2011).

Interessanterweise gibt es auch Befunde, die zeigen, dass Personen häufiger speziell auf die Alarmstufe von LAS reagieren als auf die Alarmstufe von BAS, selbst wenn die Reliabilität der Stufen dieselbe ist (Gérard, 2011; Wiczorek, 2012). Eine Erklärung ist, dass die Alarmstufe des LAS subjektiv als reliabler und verlässlicher wahrgenommen wird. Möglicherweise wird dies durch einen Kontrasteffekt hervorgerufen, der durch die Existenz einer Warnstufe im LAS erzeugt wird: Da die Warnstufe nur einen geringen PPV besitzt, könnten die Operateure die Reliabilität

der Alarmstufe als größer wahrnehmen, als sie in Wirklichkeit ist, und ihr folglich mehr vertrauen als der Alarmstufe des BAS.

Fragestellung

Das vorliegende Experiment knüpft an diese Überlegung an und überprüft, ob die Alarmstufen zweier Alarmsysteme mit exakt derselben Reliabilität subjektiv unterschiedlich bewertet werden in Abhängigkeit davon, ob das Alarmsystem zusätzlich eine Warnstufe besitzt oder nicht. Als subjektive Bewertung werden dabei zum einen die wahrgenommene Reliabilität der Alarmstufe und zum anderen das Vertrauen in die Alarmstufe betrachtet. Daraus ergeben sich die folgenden beiden Hypothesen:

1. Personen schätzen die Reliabilität der Alarmstufe des LAS größer ein als die Reliabilität der Alarmstufe des BAS.
2. Personen vertrauen der Alarmstufe des LAS mehr als der Alarmstufe des BAS.

Methode

An der Studie haben 29 Personen (17 weibliche und 12 männliche) im Alter von 18 bis 40 Jahren ($M = 25.29$, $SD = 2.89$) teilgenommen. Als Aufwandsentschädigung erhielten die Probanden 11 Euro und je nach Leistung im Experiment einen zusätzlichen Betrag in Höhe von 7 Euro.

Zur Simulation der Versuchsumgebung wurde das Programm M-TOPS verwendet. Das Ziel war es, typische Anforderungen von Multitasking Aufgaben an Operateure in Kontrollräumen der Prozessindustrie in einfacher Weise nachzustellen. Die Teilnehmer sollten dabei drei Aufgaben gleichzeitig bearbeiten. In einer der Aufgaben, bei der sie beurteilen sollten, ob das chemische Endprodukt ein korrektes Molekulargewicht besitzt, wurden sie zusätzlich durch ein Alarmsystem unterstützt, das folgendermaßen funktionierte: Durch ein grünes Licht und eine entsprechende Textanzeige gab das Alarmsystem an, dass das Molekulargewicht in Ordnung ist. Ein rotes Licht ging einher mit einer Anzeige, dass das Gewicht zu hoch ist. Die zusätzliche Warnstufe des LAS, die aussagte, dass das Molekulargewicht eventuell zu hoch ist, wurde durch ein oranges Licht angezeigt. Der PPV der Alarmstufen beider Alarmsysteme lag bei 0.6. Die Warnstufe des LAS hatte einen PPV von 0.18. Die unterschiedlichen PPVs wurden lediglich durch eine Veränderung der Base Rate des kritischen Ereignisses für beide Systeme erreicht.

Design

Das Experiment basiert auf einem 2 (Alarmsystem) x 2 (Abfolge) Versuchsdesign. Das Alarmsystem wurde als Innersubjektfaktor in den Versuch aufgenommen und hatte die beiden Stufen BAS und LAS. Die Abfolge besaß ebenfalls zwei Stufen:

Entweder haben die Probanden am ersten Tag mit dem BAS gearbeitet und am zweiten Tag mit dem LAS oder umgekehrt. Dieser Faktor wurde als Zwischensubjektfaktor realisiert. Die Probanden wurden randomisiert einer der beiden Gruppen zugeteilt.

Die abhängigen Variablen waren zum einen die wahrgenommene Reliabilität der Alarmstufe und zum anderen das Vertrauen in die Alarmstufe. Zur Erfassung der wahrgenommenen Reliabilität wurden die Personen direkt nach einer Einschätzung des PPV gefragt. Dazu sollten sie ein Kreuz auf einer 10cm langen Linie setzen, um anzugeben, in wie viel Prozent der Fälle die Alarmstufe ihrer Einschätzung nach tatsächlich mit einem kritischen Ereignis verbunden war, also einen Hit geliefert hat. Die Enden der Linie waren dabei mit „0% korrekt“ und „100% korrekt“ beschriftet. Als Messwert diente jeweils die Distanz vom Beginn der Linie bis zum Kreuz in Millimetern.

Das Vertrauen in die Alarmstufe wurde analog dazu ebenfalls durch ein einzelnes Item und eine 10cm lange Linie erhoben. Die Enden der Linie waren in diesem Fall mit „kein Vertrauen“ und „starkes Vertrauen“ beschriftet. Zusätzlich wurde den Probanden der Fragebogen von Wiczorek (2011) vorgelegt, der sich auf ihr Vertrauen zum jeweiligen Alarmsystem insgesamt bezog und aus 16 Items bestand.

Ablauf

Zu Beginn sollten die Teilnehmer eine Einverständniserklärung unterschreiben. Danach lasen sie am Bildschirm die Instruktionen zum Versuch und hatten die Möglichkeit, die einzelnen Aufgaben des Versuchs getrennt zu üben. Anschließend wurde den Versuchspersonen erklärt, dass sie in einer der Aufgaben durch ein Alarmsystem unterstützt werden, das jedoch nicht immer fehlerfrei arbeitet. Im Rahmen eines Trainingsdurchgangs mit 50 Trials, bei dem die Probanden eine akustische Rückmeldung über ihre Leistung erhielten, hatten sie die Möglichkeit, einen Eindruck von der Reliabilität des Alarms zu bekommen. Ihre wahrgenommene Reliabilität sollten sie nach dem Trainingsdurchgang getrennt für jede Stufe des Alarmsystems angeben.

Im Anschluss begann der eigentliche Experimentaldurchgang, in dem die Leistung der Versuchsteilnehmer aufgezeichnet wurde. Diese wurde quantifiziert, indem die Probanden Punkte gutgeschrieben bekamen für jede gelöste Parallelaufgabe sowie einen Punktabzug für falsche Reaktionen auf einen Alarm erhielten, das heißt für Misses und False Alarms. Abschließend hatten die Probanden die Aufgabe, ihr Vertrauen in die einzelnen Stufen des Alarmsystems anzugeben und den Fragebogen zum Vertrauen insgesamt auszufüllen.

Nach mindestens 24 und höchstens 48 Stunden erschienen die Personen dann zum zweiten Teil des Versuchs. Die Vorgehensweise am zweiten Termin war

äquivalent zum ersten Tag, allerdings wurde diesmal das jeweils andere Alarmsystem eingesetzt.

Ergebnisse

Bis auf eine Person haben alle Probanden vollständig am Experiment teilgenommen.

Eine Fragestellung der Studie war herauszufinden, ob die Reliabilität der Alarmstufe des LAS größer wahrgenommen wird als die des BAS. Die Ergebnisse liefern hier keine Bestätigung. Die Teilnehmer beurteilten die Alarmstufe des LAS ($M = 60.54$, $SD = 17.93$) nicht als signifikant reliabler als die Alarmstufe des BAS ($M = 60.86$, $SD = 18.65$), $F(1,27) = 1.620$, $p = .21$.

Als zweites sollte untersucht werden, ob die Personen der Alarmstufe des LAS mehr vertrauen als der Alarmstufe des BAS. Da ein Proband die Fragen zum Vertrauen nicht beantwortete, musste er aus der Auswertung herausgenommen werden. Die Ergebnisse zeigen einen signifikanten Unterschied in der Beurteilung des Vertrauens abhängig von der Art des Alarmsystems, ($F(1,26) = 10.732$, $p < .05$). Die Probanden vertrauten der Alarmstufe des LAS mehr ($M = 59.50$, $SD = 23.85$) als der des BAS ($M = 42.11$, $SD = 23.97$).

Zusätzlich ergab sich ein Interaktionseffekt zwischen dem Alarmsystem und der Abfolge von $F(1, 26) = 10.13$, $p < .05$. Am ersten Tag wurde das Vertrauen in die Alarmstufen beider Systeme nicht signifikant unterschiedlich angegeben: Für das LAS ergab sich ein Mittelwert von $M = 52.80$ ($SD = 29.12$), für das BAS hingegen lag dieser bei $M = 50.67$ ($SD = 20.97$). Am zweiten Tag zeigten sich jedoch deutliche Unterschiede. Für das Vertrauen in die Alarmstufe des LAS ergab sich ein Mittelwert von $M = 67.23$ ($SD = 13.03$), der höher liegt als der Wert für das BAS am ersten Tag. Der Mittelwert für das Vertrauen beim BAS lag am zweiten Tag bei $M = 35.25$ ($SD = 24.68$) und damit niedriger als der Wert für das LAS am ersten Tag.

Ein signifikanter Effekt des Alarmtyps auf das generelle Vertrauen in das System konnte nicht gefunden werden. Die Bewertungen der beiden Alarmsysteme auf dem Fragebogen zum generellen Vertrauen unterschieden sich nicht signifikant, ($F(1,26) = 2.99$, $p = .095$).

Diskussion

Die Ergebnisse unterstützen die aufgestellten Hypothesen teilweise. Die Probanden beurteilten die Alarmstufe des LAS zwar nicht als reliabler als die Alarmstufe des BAS, vertrauten ihr aber dennoch mehr.

Die gefundenen Unterschiede im Vertrauen scheinen sich allerdings nicht auf einen Kontrasteffekt durch die Warnstufe des LAS zurückführen zu lassen, wie er eingangs beschrieben wurde. In diesem Fall hätten die Versuchspersonen nämlich generell ein größeres Vertrauen in das LAS angeben müssen, unabhängig davon, ob es am ersten oder am zweiten Tag eingesetzt wurde. Am ersten Tag stuften die Versuchspersonen ihr Vertrauen jedoch in beide Alarmsysteme gleich ein. Offenbar wurde der Unterschied des Vertrauens in die beiden Systeme also durch die Testung der beiden Alarmsysteme als Innersubjektfaktor hervorgerufen – ein Unterschied im wahrgenommenen Vertrauen lässt sich nur im Vergleich mit der Bewertung am zweiten Tag finden.

Darüber hinaus wäre noch ein weiterer Effekt denkbar, der einen Einfluss auf das empfundene Vertrauen der Probanden in die beiden Alarmsysteme haben könnte. LAS vermitteln konkrete und konsistente Informationen über die Zuverlässigkeit ihrer Diagnose. Es konnte gezeigt werden, dass das Vertrauen in Automationsysteme größer ist, wenn sie Informationen liefern, die konkrete und konsistente Details enthalten (Lee & See, 2004). Darüber hinaus fanden Antifakos, Kern, Schiele und Schwaninger (2005), dass das Vertrauen von Personen in ein System steigt, wenn Informationen über die Sicherheit von Diagnosen des Systems gegeben wird. Auf Grundlage dieser Befunde scheint es plausibel, dass das größere Vertrauen in LAS dadurch verursacht wird, dass dieses dem Operateur detailliertere Informationen zur Verfügung stellt. Problematisch an dieser Interpretation ist allerdings, dass sich ein solcher Effekt auch im Vertrauen der Probanden zum Alarmsystem insgesamt widerspiegeln sollte. Die Ratings der Personen auf den Fragebögen zum Vertrauen insgesamt unterscheiden sich jedoch nicht signifikant zwischen BAS und LAS.

Auffällig im Zusammenhang mit den nicht signifikanten Ergebnissen bei der Einschätzung der Reliabilität der Alarmstufen ist im Übrigen, dass die Versuchspersonen in der Lage waren, sehr präzise Schätzungen des tatsächlichen PPV beider Alarmsysteme abzugeben. Die Versuchspersonen konnten also offenbar zwar die Alarmstufen beider Systeme rational korrekt einstufen, jedoch wurden die Alarmstufen subjektiv nicht gleich bewertet.

In der vorliegenden Studie konnte gezeigt werden, dass Personen ein größeres Vertrauen in die Alarmstufe eines LAS empfinden als in die eines BAS. Erste Überlegungen über die dahinterliegenden Ursachen wurden hier bereits angestellt – abschließend konnte aber nicht geklärt werden, wie das unterschiedliche Vertrauen zustande kam. Wie Lee & See (2004) jedoch ausführen, spielt Vertrauen in Automation eine entscheidende Rolle bei der Vorhersage menschlichen Verhaltens. In Zukunft sollte daher besonderes Augenmerk auf den Aspekt des Vertrauens gelegt und untersucht werden, welche Eigenschaften von LAS hierauf einen Einfluss ausüben.

Literatur

- Antifakros, S., Kern, N., Schiele, B., & Schwaninger, A. (2005). *Towards Improving Trust in Context-Aware Systems by Displaying System Confidence*. Präsentiert auf MobileHCI '05: Proceedings of the 7th international conference on Human Computer Interaction with mobile devices & services, New York (S. 9 – 14). New York: ACM.
- Bliss, J. P., Gilson, R. D., & Deaton, J. E. (1995). Human probability matching behavior in response to alarms of varying reliability. *Ergonomics*, 38(11), 2300–2312.
- Bustamante, E. A., & Bliss, J.P.(2005). *Effects of workload and likelihood information on human response to alarm signals*. Präsentiert auf Proceedings of the 13th International Symposium on Aviation Psychology, Oklahoma City, OK: Wright State University, 81–85.
- Bustamante, E. A. (2008). Implementing likelihood alarm technology in integrated aviation displays for enhancing decision-making: A two-stage signal detection modeling approach. *International Journal of Applied Aviation Studies*, 8(2), 241–261.
- Breznitz, S. (1984). *Cry wolf: the psychology of false alarms*. Hillsdale N.J.: Lawrence Erlbaum Associates.
- Gérard, N. (2012). Verhaltenseffektivität von Alarmen: Experimentelle Untersuchungen zum Einfluss von Reliabilität und Prüfmöglichkeit auf die Anwendung von Heuristiken (Dissertation). Technische Universität Berlin, Deutschland. Abgerufen von <http://opus.kobv.de/tuberlin/volltexte/2012/3411/>
- Getty, D., Swets, J. A., Pickett, R. M., & Gonthier, D. (1995). System operator response to warnings of danger: A laboratory investigation of the effects of the predictive value of a warning on human response time. *Journal of Experimental Psychology: Applied*, 1(1), 19–33.
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1), 50–80.
- Madhavan, P., Wiegmann, D. A., & Lacson, F. C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(2), 241–256.
- Ragsdale, A., Lew, R., Dyre, B., Boring, R., & Gertman, D. (2012). *Alarm Strategy and Complexity: Predictions of Operator Response*. Präsentiert auf Nuclear Power Instrumentation and Control and Human Machine Interface Technology, San Diego. Idaho National Laboratory (INL).
- Swets, J. A. (1964). *Signal detection and recognition by human observers*. New York: Wiley.
- Swets, J. A. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *American Psychologist*, 47(4), 522–532.

- Wiczorek, R. (2011). Entwicklung und Evaluation eines mehrdimensionalen Fragebogens zur Messung von Vertrauen in technische Systeme. In S. Schmid, M. Elepfandt, J. Adenauer, & A. Lichtenstein (Hrsg.), *Reflexionen und Visionen der Mensch-Maschine-Interaktion – Aus der Vergangenheit lernen, Zukunft gestalten*, 9. Berliner Werkstatt Mensch-Maschine-Systeme (Bd. 33, S. 621–626). Berlin: VDI.
- Wiczorek, R., & Manzey, D. (2011). Evaluating likelihood alarm systems as an alternative to binary alarm systems. In D. Waard, N. Gérard, L. Onnasch, R. Wiczorek, & D. Manzey (Hrsg.), *Human centred automation* (S. 69–83). Maastricht, NL: Shaker Publishing.
- Wiczorek, R. (2012). *Verhaltenswirksamkeit von Likelihood-Alarmsystemen* (Dissertation). Technische Universität Berlin, Deutschland.

Annotationen für CAD-Modelle in virtueller Umgebung

Elisabeth Dittrich¹

Schlüsselwörter: Virtual Reality, CAVE, CAD, Information Visualization

Zusammenfassung

Die Validierung von CAD Modellen in begehbaren virtuellen Umgebungen ist sehr effizient, da die dreidimensionale Ansicht in Originalgröße Vorteile gegenüber der Desktopansicht bringt. Zu einem ganzheitlichen Review gehören jedoch auch Metainformationen bezüglich des Modells, welche bisher kaum in virtuellen Umgebungen integriert sind. Dass dies jedoch notwendig ist und beispielsweise durch strukturierte Annotationen umgesetzt werden kann zeigt dieser Artikel im Überblick.

Motivation

Bei der Produktentwicklung von mechatronischen Systemen, zum Beispiel Fahrzeugen, werden Visualisierungen benötigt, die den Überblick über die komplexen und umfangreichen Modelle bereitstellen. Diese visuelle Validierung kann in immersiven virtuellen Umgebungen durchgeführt werden, bei denen das Fahrzeugmodell in Originalgröße erlebt werden kann. Bei einem virtuellen Design Review in einer CAVE (= Cave Automatic Virtual Environment) oder vor einer Powerwall wird das Modell angezeigt, und man kann ggf. mit ihm interagieren (verschieben, skalieren, Schnitte einfügen), aber auch textuelle Informationen hinzufügen, bspw. Längenangaben zwischen zwei Maßpunkten. In Zukunft werden die Menge der abgebildeten Informationen in virtuellen Umgebungen steigen, um die Effektivität der Validierung in einer Virtuellen Realität zu steigern. Das Anordnen dieser Informationen in s.g. Information Rich Virtual Environments (IRVEs, Polys & Bowman, 2004) wurde in den Bereichen Informatik und Informationsvisualisierung untersucht (Polys & Bowman, 2004; Polys, Kim & Bowman, 2005; Polys, Bowman & North, 2010).

Ziel

Diese Promotionsarbeit trägt dazu bei, die steigende Komplexität der Darstellung von Produktdaten in Virtuellen Umgebungen (in diesem Fall einer CAVE) nutzergerecht zu gestalten. Dabei werden u. a. Methoden des Usability Engineering, der experimentellen Psychologie und der Arbeitswissenschaft verwendet. Ziel ist es, einen funktionalen Demonstrator zu entwickeln, welcher die Gestaltungsrichtlinien prototypisch umsetzt.

¹ Graduiertenkolleg prometei, TU Berlin, Fraunhofer IPK

Methode und Ergebnisse

Die Gestaltungsrichtlinien wurden durch die folgenden vier Untersuchungen erforscht:

1) Nutzer- und Kontextanalysen zum Umgang mit komplexen Produktstrukturen und virtuellen Umgebungen: Durchgeführt wurde eine Beobachtung von Design Reviews bei großen Automobilherstellern, bei welcher mehrere CAD-Modelle in einer CAVE verglichen wurden. Der Output dieser Untersuchung war eine Aufstellung prozessrelevanter Komponenten im Sinne der MTO (Mensch Technik Organisation) - Analyse (Dunckel, 1999). Folgendes Schemata (Abb.1) ist dabei entstanden:

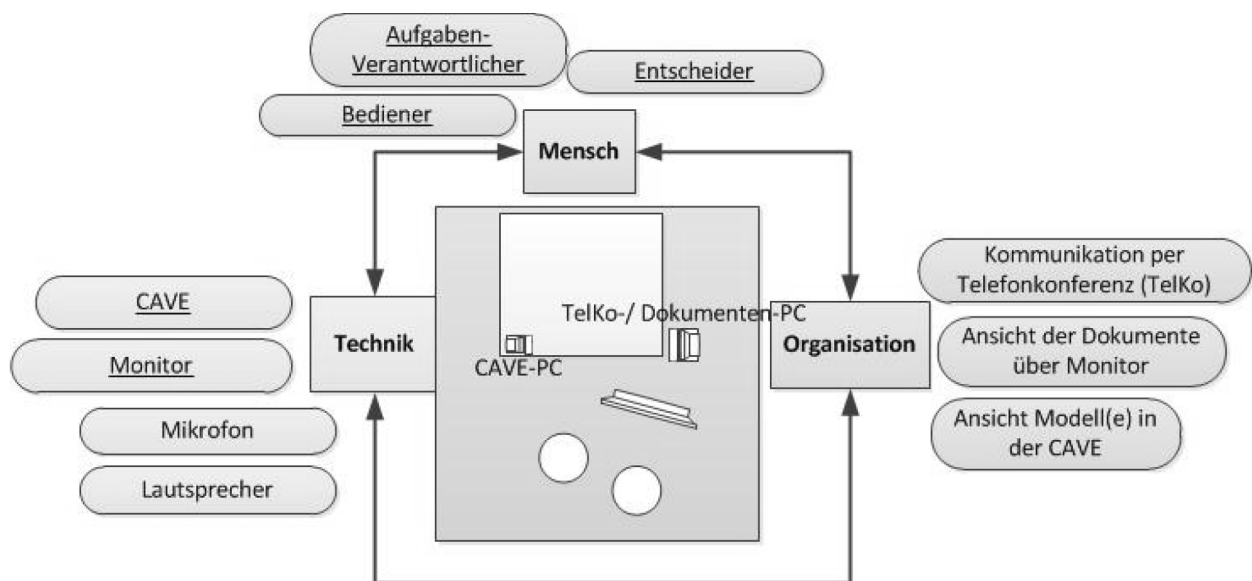


Abb.1: MTO-Analyse eines beobachteten Design Reviews in einer CAVE.

Bezüglich der Technik konnte ein komplexes Zusammenspiel verschiedener Geräte festgestellt werden. In Bezug auf diese Arbeit besteht die Erkenntnis darin, dass eine CAVE nicht als selbständig funktionierendes Arbeitsumfeld funktioniert. Aus der Vielzahl an Technik ergibt sich die Notwendigkeit von Technik-Bedienern, welche den CAVE-PC und den Dokumenten-PC unabhängig voneinander handhaben. Festzuhalten ist also hier, dass beim Gebrauch einer CAVE nicht nur die Anwender, sondern auch Operateure notwendig sind. In Bezug auf die wurde beobachtet, dass mehrere Medien abwechselnd verwendet wurden, was zu einer ständigen Bewegung der Personen führte und außerdem zu getrennten Ansichten von in Bezug stehenden Informationen. Auch hier ist der Schluss in Bezug auf die CAVE, dass sie als alleiniges Arbeitswerkzeug für ein Design Review nicht ausreicht. Fazit ist, dass die CAVE mehr anbieten sollte, als nur ein Abbild (das CAD-Modell), da weitere Informationen bei einem Design Review benötigt werden.

2) Laboruntersuchung zur Lesbarkeit von Text in virtuellen Umgebungen (Dittich, Brandenburg und Beckmann-Dobrev, 2013): Durch die Notwendigkeit von

textlichen Informationen in virtueller Umgebung stellte sich die Frage nach der Lesbarkeit und die damit verbundene Schriftauflösung und -größe der Buchstaben, um Lesbarkeit zu gewährleisten.

Die Ermittlung der Lesbarkeitsschwellen von "normal" projizierten und stereoskopisch projizierten Buchstaben ergab einen tendenziell signifikanten Unterschied. Fazit ist hierbei, dass Schrift in virtuellen Umgebungen angelehnt an zweidimensionale Gestaltungsrichtlinien (z. B. DINI 9241-303, 2008) formatiert werden kann, wenn diese an die Auflösungskapazität der Beamer angepasst werden. Diese Anpassung sollte sich mindestens auf einen Pixel mehr, bei 1mm Pixelgröße, belaufen.

3) Explorative Untersuchung zur nutzerzentrierten Darstellung von Produktstrukturen (Dittrich & Israel, 2012): Das Anordnen von Begriffen ist im dreidimensionalen Raum weit umfangreicher, als beispielsweise auf einer Präsentationsfolie (zweidimensionale Fläche). Zum Generieren vielseitiger Anordnungsmöglichkeiten arrangierten die Versuchspersonen eine feste Anzahl von Produktdaten (Abbildung 2), wobei ihnen zeitlich und gestalterisch keinerlei Einschränkungen gemacht wurden.



Abb.2.: Versuchsperson beim freien Arrangieren von Produktdaten in der CAVE.

Das Ergebnis liegt zusammengefasst darin, dass ein Modell welches Annotationen verwendet, aber auch zusätzlich die hierarchische Struktur zwischen den Begriffen aufgreift, von den Nutzern am häufigsten erstellt wurde.

4) Laboruntersuchung zur Effizienz von Visualisierungen von Modell- und Strukturrelationen: Ein wesentlicher Nachteil von Annotationen ist das Fehlen von strukturellen Informationen, beispielsweise Zusammenhänge zwischen den Begriffen. Um zu prüfen ob dieser Nachteil gegenüber der herkömmlichen hierarchischen Darstellung in diesem Kontext auf die Effizienz auswirkt, wurde eine Untersuchung mit dem Paradigma der visuellen Suche durchgeführt.

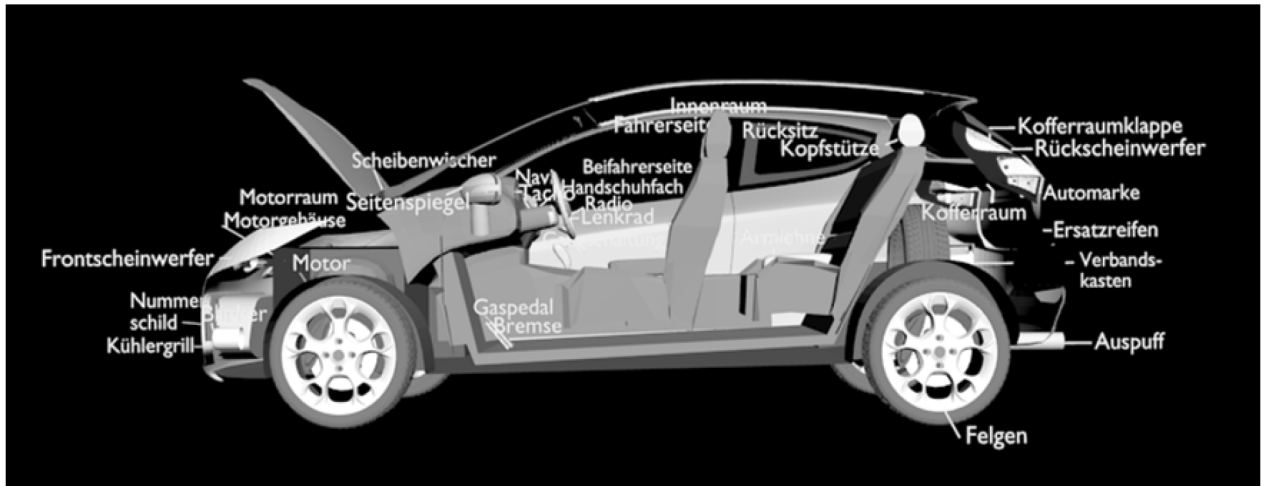


Abb. 3: Stimulusmaterial eines annotierten Fahrzeugs als stereoskopisches Modell in der CAVE.

Zusammenfassend schnitt im Vergleich die annotierte Variante (Abbildung 3) gegenüber der hierarchischen Struktur signifikant effizienter ab.

Fazit und Ausblick

Im Vordergrund der Gestaltung von Informationen in einer immersiven virtuellen Umgebung steht vor allem der Nutzer. Er sollte interagieren und individuell die Visualisierung einrichten können. Des Weiteren führt die Form der Annotation zu effizienteren Sucherfolgen. Das Formulieren der Gestaltungsrichtlinien auf Basis der empirischen Erkenntnisse steht zurzeit noch aus.

Durch das intelligente Anreichern des CAD-Modells durch Metainformationen wird das virtuelle Validieren in Design Reviews attraktiver und somit der Arbeitsprozess effizienter.

Literatur

- DIN EN ISO 9241(2008). Ergonomie der Mensch-System-Interaktion – Teil 303: Anforderungen an elektronische optische Anzeigen. Beuth Verlag GmbH, Berlin
- Dittrich, E., & Israel, J. H. Arrangement of Product Data in CAVE systems. <http://www.cse.ohio-state.edu/~raghu/teaching/CSE5544/Visweek2012/scivis/posters/dittrich.pdf>.

- Dittrich, E., Brandenburg, S., & Beckmann-Dobrev, B. (2013). Legibility of Letters in Reality, 2D and 3D Projection. In *Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments* (pp. 149-158). Springer Berlin Heidelberg.
- Dunckel, H. (Ed.). (1999). *Handbuch psychologischer Arbeitsanalyseverfahren* (Vol. 14). vdf Hochschulverlag AG
- Polys, N. & Bowman, D. (2004). Design and display of enhancing information in desktop information-rich environments: challenges and techniques. *Virtual Reality* 8, pp. 41-54
- Polys, N. Kim, S. & Bowman, D. (2005). Effects of Information Layout, Screen Size, and Field of View on User Performance in Information-Rich Virtual Environments. Proc. ACM Symp. *Virtual Reality Software and Technology*, ACM Press, pp. 46-55
- Polys, N., Bowman, D. & North, C. (2010). The role of Depth and Gestalt cues in information-rich virtual environments. *International Journal of Human-Computer Studies*, 69, S. 30-51

Aktivitätsmessung in der Orthesenversorgung der unteren Extremitäten

Laura Doria¹, Susanne Dannehl² und Marc Kraft²

Schlüsselwörter: Aktivitätsmessung, Orthese, Aktivitätsintensität, Monitoring-System, Trageverhalten

Zusammenfassung

Im Rahmen der Entwicklung eines Multi-Sensor-Monitoring-System zur Erfassung des Trageverhaltens von Patienten mit Orthesen der unteren Extremitäten im Forschungsschwerpunkt „Prospektiv-nutzergerechte Gestaltung von Medizinprodukten“ wurde die Möglichkeit untersucht, Aktivitätsmessungen mit einem dreiachsigen Beschleunigungssensor durchzuführen.

An dem Experiment nahmen 31 Probanden teil, die instrumentierte Orthesen unterschiedlicher Hersteller während Bewegungs- und Ruhephasen trugen.

Die Beschleunigungsdaten wurden anhand der Aktivitätsintensität (Signal Magnitude Area, SMA) verarbeitet und statistisch ausgewertet. Dabei wurde ersichtlich, dass die SMA bei Bewegung signifikant höher ist als in Ruhe. Die Berechnung einer einfaktoriellen ANOVA mit den abhängigen Variablen „Ruhe“ und „Bewegung“ und dem Faktor „Orthesentyp“ zeigte signifikante Orthesenspezifische Unterschiede in der Variable „Bewegung“.

Die Bewertung der bestmöglichen SMA-Grenzwerte für die Erkennung von Bewegungs- und Ruhephasen ergab die Notwendigkeit der Nutzung von angepassten Grenzwerten für die zwei Hilfsmittelkategorien „Sprunggelenk-Fuß-Orthesen“ und „funktionelle Knieorthesen“.

Hintergrund

Im Forschungsschwerpunkt „Prospektiv-nutzergerechte Gestaltung von Medizinprodukten“ wird ein eHealth-Konzept zur Akzeptanzverbesserung der Orthesenversorgung entwickelt. Das Konzept besteht aus einem Multi-Sensor-Monitoring-System zur Erfassung des Trageverhaltens von Patienten im Umgang mit Orthesen der unteren Extremitäten und aus einem Online-Portal zum Informations- und Datenaustausch (Doria, Dannehl & Kraft, 2013).

Das Messkonzept besteht aus einem dreiachsigen Beschleunigungssensor und zwei Feuchte- und Temperatursensoren. Die Tragezeiten werden mittels der Temperaturdifferenz zwischen Innen- und Außenseite der Orthese berechnet. Die mikroklimatischen Eigenschaften zwischen Orthese und Haut werden durch die An-

¹ Technische Universität Berlin, Graduiertenkolleg prometei

² Technische Universität Berlin, Fachgebiet Medizintechnik

wendung des Feuchte- und Temperatursensors an der Innenseite der Orthese erfasst. Die Erfassung der Bewegung der Patienten während der Tragezeiten wird vom Beschleunigungssensor gewährleistet (Doria et al., 2013).

Die Aktivitätserkennung aus Beschleunigungsdaten ist hierbei komplexer als die Auswertung der Daten hinsichtlich Tragezeiten und Mikroklima. Ziel dieser Untersuchung war daher, einen schlanken Algorithmus zu identifizieren, um Ruhe- und Bewegungsphasen von Probanden im Umgang mit einer Orthese zu erkennen.

Methoden

Die technische Umsetzung erfolgte durch den kommerziellen Datenlogger MSR145 der Firma MSR Electronics GmbH.

Der externe Drei-Achsen-Beschleunigungssensor (Messbereich= ± 15 g (-20...+65 °C); Messgenauigkeit= $\pm 0,15$ g (+25 °C); Messfrequenz=50 Hz) wurde an drei unterschiedlichen Orthesen befestigt:

- Sprunggelenk-Fuß-Orthese (Fußorthese): Unterschenkel, ventral;
- erste funktionelle Knieorthese (Knieorthese 1): Unterschenkel, ventral;
- zweite funktionelle Knieorthese (Knieorthese 2): Unterschenkel, dorsal.

Für das Laborexperiment konnten 31 Probanden gewonnen werden, die in die Gruppen „Fußorthese“, „Knieorthese 1“ und „Knieorthese 2“ eingeteilt wurden.

Nachdem sie die Orthese angezogen hatten, wurden sie gebeten, unterschiedliche Aktivitäten durchzuführen. Diese können in die zwei Kategorien „Ruhe“ (Sitzen, Stehen) und „Bewegung“ (Laufen, Treppen steigen, Treppen runter laufen) unterteilt werden können, wobei die Geschwindigkeit und Dauer der Aktivitäten nicht vorgegeben wurde.

Die erhobenen Signale für die drei Beschleunigungsachsen wurden mittels eines gleitenden Mittelwertes der dritten Ordnung gefiltert, um die Wirkung des Rauschens zu reduzieren.

Die gemessene Beschleunigung besteht aus zwei Komponenten. Die erste ist auf die Erdbeschleunigung und die zweite ist auf die menschliche Bewegung zurückzuführen. Da für die Aktivitätserkennung nur die zweite Komponente von Bedeutung ist, wurden die Signale mit einem Tiefpassfilter mit unendlicher Impulsantwort verarbeitet (Durchlassbereich=0,01dB; Sperrbereich=100dB; normierte Grenzfrequenz=0,28). Die gesuchte Komponente ist die Differenz zwischen dem ursprünglichen und dem gefilterten, von der Erdbeschleunigung verursachten Signal.

Das metabolische Äquivalent (Energy Expenditure, EE) ist proportional zu der Aktivitätsintensität (Bouten, Koekkoek, Verduin, Kodde & Janssen, 1997), welche folgendermaßen aus der Beschleunigung berechnet werden kann:

$$\text{Signal Magnitude Area (SMA)} = \frac{1}{k} * \left(\sum_{i=1}^k |x_i| + \sum_{i=1}^k |y_i| + \sum_{i=1}^k |z_i| \right),$$

wobei x_i , y_i und z_i die Beschleunigungswerte und k die Größe des Zeitfensters sind (Krassnig, Tantinger, Hofmann, Wittenberg & Struck, 2010).

Die Aktivitätsintensität wurde schon in mehreren Studien verwendet, um Bewegungs- und Ruhephasen während der Messung mit Beschleunigungssensoren zu untersuchen (bspw. Krassnig et al., 2010; Karantonis, Narayanan, Mathie, Lovell & Celler, 2006; Khan, Lee, Lee & Kim, 2010; Mathie, Coster, Lovell & Celler, 2003). Dieser Ansatz benötigt darüber hinaus keine Implementierung aufwändiger Auswertungsalgorithmen und wurde daher hier weiter verfolgt.

Ergebnisse

In IBM SPSS Statistics 21 wurden die Mittelwerte für die SMA pro Minute bei Bewegung und in Ruhe für die Gesamtstichprobe und für jede Orthese einzeln durch vier Student T-Tests für verbundene Stichproben ausgewertet. Die SMA für die Gesamtstichprobe in Ruhe (Mittelwert=0,0179g; SD=0,0044g) war signifikant niedriger als die SMA bei Bewegung (Mittelwert=0,1490g; SD=0,0430; $t(30)=17,421$; $p<0,001$). Ähnlich verhielten sich die Mittelwerte für die einzelnen Orthesen:

- Fußorthese: Mittelwert in Ruhe=0,0158g; SD=0,0052g; Mittelwert bei Bewegung=0,1170g; SD=0,0349g; $t(10)=10,054$; $p<0,001$;
- Knieorthese 1: Mittelwert in Ruhe=0,0186g; SD=0,0035g; Mittelwert bei Bewegung=0,1483g; SD=0,0209g; $t(9)=20,098$; $p<0,001$;
- Knieorthese 2: Mittelwert in Ruhe=0,0195g; SD=0,0037g; Mittelwert bei Bewegung=0,1848g; SD=0,0412; $t(9)=12,338$; $p<0,001$.

Die Berechnung einer einfaktoriellen ANOVA mit den abhängigen Variablen „Ruhe“ und „Bewegung“ und dem Faktor „Orthesentyp“ zeigte signifikante Unterschiede in der Variable „Bewegung“ ($F(2,28)=10,782$, $p<0,001$, $\eta_p^2=0,435$). Aus den Post-Hoc-Tests (Scheffé) wurde deutlich, dass die Fußorthese und die Knieorthese 2 sich in der Variable „Bewegung“ ($p<0,001$) signifikant unterschieden.

In der Abbildung 1 sind die möglichen Grenzwerte (x-Achse) zur Unterscheidung von Ruhe- und Bewegungsphasen dargestellt. Ein Fehler ergibt sich, wenn bei einem Proband die Bewegungszeit nicht vollständig korrekt erkannt wurde. Der prozentuale Fehler (y-Achse) wurde wie folgt berechnet:

$$\text{Fehler [\%]} = \frac{\text{Fehleranzahl}}{\text{Probandenanzahl}} * 100.$$

Die Grenzwerte, bei denen Minima in den Fehlern auftreten, sind in der Tabelle 1 zusammengefasst.

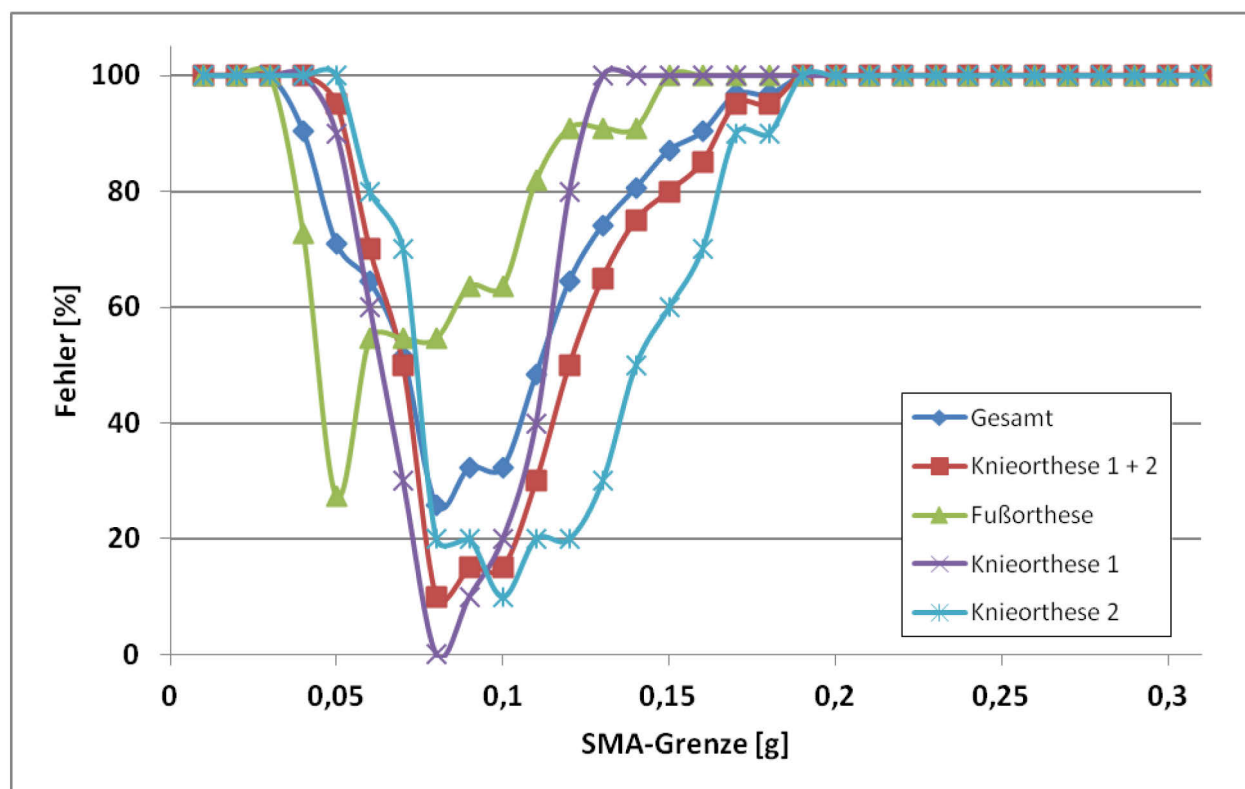


Abb.1: Diagramm der berechneten Fehlerfälle in Funktion der SMA-Grenzwerten zur Unterscheidung von Ruhe und Bewegung.

Tab.1: SMA-Grenzwerte und Fehler (Best Case).

	SMA-Grenzwert [g]	Fehler [%] (N. Fehler/ N.Probanden)
Gesamt	0,08	26 (8/31)
Knieorthese 1 + 2	0,08	10 (2/20)
Fußorthese	0,05	27 (3/11)
Knieorthese 1	0,08	0 (0/10)
Knieorthese 2	0,10	10 (1/10)

Diskussion

Die SMA erwies sich als ein geeignetes Auswerteverfahren für die Erkennung von Bewegungs- und Ruhephasen bei der Nutzung von Orthesen der unteren Extremitäten.

täten. Der implementierte Algorithmus konnte die vorhandene Datenmenge innerhalb von wenigen Minuten bearbeiten.

Die Nutzung eines einzigen Grenzwertes führte zu einem Fehler von 26%. Dies ist darauf zurückzuführen, dass die Fußorthese und die Knieorthese 2 sich in der Variable „Bewegung“ signifikant unterschieden. Eine Verbesserungsmöglichkeit ist hier die Einführung von zwei angepassten Grenzwerten für die Kategorien „Sprunggelenk-Fuß-Orthesen“ und „funktionelle Knieorthesen“. Auf diese Weise kann der Fehler auf 16% gesenkt werden. Diese Kategorisierung kann auch im Rahmen einer Studie an Patienten mit mehreren Orthesen unterschiedlicher Hersteller übernommen werden. Obwohl die Anwendung von einem Grenzwert pro Orthesentyp (bzw. Orthesenmarke) den Fehler auf 10% senken würde, ist dies im Rahmen größerer Studien aufgrund der großen Anzahl an Orthesenmodelle auf dem Markt kaum zu gewährleisten.

Bei dieser Bewertung wurde einen Proband als Fehlerfall berechnet, selbst wenn die Bewegungszeit um eine einzige Minute über- oder unterschätzt wurde. In weiteren Untersuchungen sollte die Präzision des Auswertungsverfahrens auch hinsichtlich der Fehler innerhalb der einzelnen Probandenmessungen bewertet werden.

Literatur

- Bouten, C., Koekkoek, K., Verduin, M., Kodde, R., & Janssen, J. (1997). A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Transactions on Biomedical Engineering*, 44(3), 136–147. doi:10.1109/10.554760
- Doria, L., Dannehl, S., & Kraft, M. (2013). Entwicklung eines eHealth-Konzeptes zur Akzeptanzverbesserung der Orthesenversorgung. In P. Haas, S. C. Semler, S. H. Schug, & J. Schenkel (Eds.), *Nutzung, Nutzer, Nutzen von Telematik in der Gesundheitsversorgung - eine Standortbestimmung* (1st ed., pp. 45–52). Berlin: TMF.
- Karantonis, D., Narayanan, M., Mathie, M., Lovell, N., & Celler, B. (2006). Implementation of a Real-Time Human Movement Classifier Using a Triaxial Accelerometer for Ambulatory Monitoring. *IEEE Transactions on Information Technology in Biomedicine*, 10(1), 156–167. doi:10.1109/TITB.2005.856864
- Khan, A. M., Lee, Y.-K., Lee, S., & Kim, T.-S. (2010). Accelerometer's position independent physical activity recognition system for long-term activity monitoring in the elderly. *Medical & biological engineering & computing*, 48(12), 1271–1279. doi:10.1007/s11517-010-0701-3
- Krassnig, G., Tantinger, D., Hofmann, C., Wittenberg, T., & Struck, M. (2010). User-friendly system for recognition of activities with an accelerometer. In *4th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2010. 22 - 25 March 2010, Munich, Germany ; [including workshop papers]*. Piscataway, NJ: IEEE.

Mathie, M. J., Coster, A. C. F., Lovell, N. H., & Celler, B. G. (2003). Detection of daily physical activities using a triaxial accelerometer. *Medical & Biological Engineering & Computing*, 41(3), 296–301. doi:10.1007/BF02348434

Avatare in der virtuellen Neurorehabilitation der unteren und oberen Extremitäten

Luara Ferreira dos Santos¹, Henning Schmidt², Christian Dohle^{3,4,5}, Jörg Krüger²

Schlüsselwörter: Virtuelle Realität, motorische Neurorehabilitation, Avatare, visuelles Feedback, systematisches Review

Zusammenfassung

In der motorischen Neurorehabilitation werden immer öfter Avatare als visuelle Rückkopplung in Therapieansätzen mit virtueller Realität (VR) angewendet. Befunde aus der Grundlagenwissenschaft weisen darauf hin, dass die Beobachtung virtueller Arme die Erregbarkeit des motorischen Cortexes erhöht. Der Einsatz von Avataren könnte somit für einen besseren motorischen Lerneffekt in der Therapie eingesetzt werden. In einem systematischen Review wurden Publikationen VR-basierter Interventionen mit Avatar-Feedbacks für die motorische Rehabilitation der unteren oder oberen Extremitäten erhoben. Ergebnisse eingeschlossener Studien zeigen einen positiven Effekt auf die motorische Erholung beim Einsatz von virtuellen Körpern (-teilen) als visuelle Rückkopplung. Die Daten ermöglichen jedoch kaum einen Vergleich von Feedbacktypen: der Einsatz von Avataren erfolgt in den meisten eingeschlossenen Studien nur unsystematisch und es fehlen klinische Studien, die die Art der Visualisierung grundlegend untersuchen. Dies wäre notwendig um zukünftige Interventionen zielgerichtet zu optimieren.

Hintergrund

Ein neuerer Ansatz in der motorischen Neurorehabilitation, z.B. nach Schlaganfall, ist die Anwendung von Therapien auf der Grundlage virtueller Realität (VR). Patientinnen und Patienten erhalten hierbei eine motorische Übungsaufgabe, deren Ausführung in einer virtuellen Umgebung visuell dargestellt wird. Zur Echtzeit-Bewegungserfassung können verschiedene Motion-Tracking Systeme genutzt werden, um die Zielgenauigkeit der Körperbewegungen als visuelle Rückkopplung (Feedback) darzustellen. Auch kommerzielle Videospiel-Konsolen, wie Nintendo Wii, Dance Dance Revolution, Sony Eye Toy und Xbox Kinect, haben in den letzten Jahren immer öfter Einzug in die motorische Neurorehabilitation gehalten und es gibt bereits mehrere Studien, die die Wirksamkeit dieser Systeme untersucht

¹ Graduiertenkolleg prometei, Forschungsgruppe Rehabilitationsrobotik (IPK/TU Berlin), Technische Universität Berlin

² Forschungsgruppe Rehabilitationsrobotik (IPK/TU Berlin), Abteilung Automatisierungstechnik, Fraunhofer Institut für Produktionsanlagen und Konstruktionstechnik (IPK), Berlin

³ Abteilung für Neurologische Rehabilitation, MEDIAN Klinik Berlin-Kladow

⁴ Centrum für Schlaganfallforschung Berlin, Charité Universitätsmedizin Berlin

⁵ Professur für Rehabilitationswissenschaften, Universität Potsdam

haben (für einen Überblick siehe z.B. Taylor, McCormick, Shawis, Impson, & Griffin, 2011).

Die Visualisierung dieser Feedbacks kann sehr unterschiedlich gestaltet sein: von abstrakten zwei- oder dreidimensionalen Spielen, über komplexe virtuelle Szenarien bis hin zu Avataren (Ferreira dos Santos, L., Schmidt, Krüger, & Dohle, 2013, in press). Letzteres ist die Darstellung virtueller Körper oder Körperteile als visueller Stellvertreter der ausgeführten Bewegungen. Diese finden immer öfter Anwendung, auch in der Geräte-gestützten motorischen Neurorehabilitation.

Grundlagenwissenschaftliche Befunde weisen darauf hin, dass eine höhere motorische Hirnaktivität durch die Darstellung virtueller Körper möglich ist. Die Studienlage zeigt, dass die Beobachtung einer Handbewegung die Erregbarkeit im motorischen Cortex stärker erhöht als bei der Beobachtung in Kombination mit einem Zielobjekt (Villiger, Chandrasekharan, & Welsh, 2011). Zudem konnte gezeigt werden, dass das Körperschema bei der Beobachtung eines bewegungssynchronen virtuellen Körperteils ähnlich integriert wird wie reale Bewegungen (Adamovich, August, Merians, & Tunik, 2009; Dohle et al., 2011). Visuelle Informationen von Körperbewegungen scheinen zudem einen unterstützenden Einfluss auf die motorische Ausübung der Bewegung zu haben (Altschuler, 2005). Gut untersucht ist der Spezialfall der Spiegeltherapie zur Behandlung von einseitiger Parese nach Schlaganfall, bei der mit Hilfe einer Spiegelung der nicht betroffenen Extremität die Illusion von Bewegungen der betroffenen Extremität erzeugt wird (Thieme, Mehrholz, Pohl, Behrens, & Dohle, 2012). Auch hier wirken Avatare in ganz ähnlicher Art und Weise wie „echte“ Spiegelungen (Dohle et al., 2011).

Alle Evidenzen zusammengekommen, macht die Darstellung von Avataren einen besseren Lerneffekt wahrscheinlich, der in der neurologischen Rehabilitation zur Steigerung der Effektivität von motorischen Therapien genutzt werden könnte. Durch den zunehmenden technischen Fortschritt können Körper und Extremitäten computergraphisch immer realer und mit höherer zeitlicher Synchronizität meßtechnisch erfasst und dargestellt werden. Dies ermöglicht neuartige und für das motorische Lernen potentiell förderliche Therapie-Szenarien, die bisher nicht umzusetzen waren.

Um den Aspekt der visuellen Information während der VR-basierten motorischen Rehabilitation der unteren und oberen Extremitäten näher zu untersuchen, wurde eine systematische Literaturrecherche durchgeführt. Ergebnisse von Studien mit Avatar-Feedbacks werden auf mögliche Implikationen für die motorische Rehabilitation diskutiert.

Methode

Im Zeitraum von November bis Dezember 2012 wurden die elektronischen Datenbanken MEDLINE (Ovid), AMED, EMBASE, CINAHL und PsychInfo auf VR-basierte visuelle Feedback-Interventionen für die motorische Rehabilitation durch-

sucht. Dabei wurde der Suchalgorithmus von Laver, George, Thomas, Deutsch, & Crotty (2011) angewendet. Eingeschlossen wurden Studien mit Interventionen für die oberen oder unteren Extremitäten, die an einer Patientenstichprobe oder gesunden Studienteilnehmerinnen und -teilnehmern angewendet wurde. Ausgeschlossen wurden Studien zum Training von kognitiven Funktionen und psychotherapeutischen Behandlungen.

Ergebnisse

Insgesamt entsprachen 142 Publikationen den Einschlusskriterien. Diese wurden nach ihren visuellen Rückkopplungen der motorischen Ausführung analysiert und in sechs Feedbackgruppen unterteilt (Ferreira dos Santos, L. et al., 2013, in press). Unter Avatar-Feedback wurden computergraphische Darstellungen von menschlichen Körpern oder Körperteilen eingeschlossen, bei denen die realen Körperbewegungen von einem virtuellen Körper oder Körperteil repräsentiert werden. Die Gruppe der Avatar-Feedbacks schloss 53 Publikationen ein, und ist die am stärksten vertretene Feedbackgruppe dieses Reviews.

Eine genauere Analyse der Visualisierungen des Avatar-Feedbacks weist drei Subgruppen auf: realistischer menschlicher Körper oder Körperteil (N= 40), gespiegelter Körper oder Körperteil (N= 6) und grobe Figur (z.B. Mii-Avatar bei Nintendo Wii; N= 9). Es zeigt sich eine sehr viel höhere Anzahl von Avatar-Feedback Publikationen für die oberen Extremitäten (OE; N= 48) als für die unteren Extremitäten (UE; N= 5) und doppelt so viele Publikationen mit einer alleinigen Anwendung von VR (N= 36) als Geräte-gestützter Interventionen in Kombination mit VR (GVR; N= 18). Es ist zu beachten, dass bei den Publikationen der oberen Extremitäten eine Publikation zwei Subgruppen („realistischer menschlicher Körper oder Körperteil“ und „grobe Figur“) und eine weitere Publikation zu VR und GVR zugeordnet wurde. Tabelle 1 stellt die Häufigkeiten entsprechend der Subgruppen dar.

Tab.1: Verteilung der Publikationen nach Avatar-Subgruppen.						
		<i>Extremitäten</i>			<i>Intervention</i>	
Avatar-Feedback		OE	UE		VR	GVR
Realistischer Körper (-teil)		37	3		22	18
Gespigelter Körper (-teil)		6	0		6	0
Grobe Figur		7	2		9	0

Die Betrachtung der Studienergebnisse zeigt, dass die Visualisierung von Avataren einen positiven Effekt auf die motorische Erholung bei einer VR-basierten Rehabilitation nach Schlaganfall haben kann (Cameirao, Bermudez i

Badia, Duarte, Frisoli, & Verschure, 2012; Connelly et al., 2010; Kam, Struzik, Jarus, & Rand, 2012).

Um überprüfen zu können, ob Avatar-Feedbacks einen besseren Effekt als andere visuelle Rückkopplungen auf die motorische Rehabilitation haben, ist es notwendig diese gegeneinander zu testen. In 13 der eingeschlossenen Publikationen werden neben einer Avatar-Darstellung auch andere Visualisierungen innerhalb derselben Studie angewandt, so dass ein Vergleich möglich wird. Jedoch können bei zwölf Studien keine aussagekräftigen Ergebnisse hierfür verwendet werden, da entweder eine unzureichende Studienqualität vorliegt (N= 6), keine Therapiestudie über einen längeren Zeitraum durchgeführt wurde (N= 2) oder die unterschiedlichen Feedbacks nicht separat analysiert wurden (N= 4). Nur die Studie von da Silva Cameirão, Bermúdez i Badia, Duarte, & Verschure (2011) kann für einen Vergleich herangezogen werden.

In dieser Studie wurde ein virtuelles motorisches Rehabilitationsspiel (Rehabilitation Gaming System, RGS) mit der Darstellung realistischer Arme und Hände aus der Egoperspektive an Schlaganfallpatienten untersucht. Während zwölf Wochen wurde eine Gruppe von zehn Patientinnen und Patienten am RGS trainiert, während zwei weitere Kontrollgruppen entweder Training an der Nintendo Wii mit VR-Spielen ohne menschlichen Darstellungen (N= 4) oder intensive Ergotherapie erhielten (N= 5). Zwischen den Kontrollgruppen zeigten sich keine signifikanten Unterschiede in den erhobenen klinischen Parametern. Die RGS-Gruppe zeigte im Inter-Gruppen-Vergleich nach dem Training signifikant bessere Ergebnisse im Fugl-Meyer Assessment Test und im Chedoke Arm and Hand Activity Inventory. Die Probanden aus dieser Gruppe hatten eine bedeutend höhere Verbesserung in ihren motorischen Fähigkeiten als die Probanden aus den Kontrollgruppen. Es ist jedoch kritisch anzumerken, dass die Ergebnisse beider Kontrollgruppen für diese Analyse zusammengefasst wurden, so dass auch hier kein idealer Vergleich zwischen verschiedenen Feedbacktypen durchgeführt wurde. Die Autoren nehmen an, dass aufgrund der Beobachtung virtueller Arme eine funktionelle Reorganisation der motorischen und prämotorischen Areale erleichtert wird (Cameirao, Bermudez i Badia, Duarte Oller, & Verschure, 2010).

Die Datenlage zeigt ein breites Spektrum an Avatar-Feedbackstudien zur motorischen VR-basierten Rehabilitation von neurologischen Patientinnen und Patienten. Unsere Ergebnisse zeigen jedoch, dass nur eine dieser Studien für einen Vergleich von Avatar-Feedback mit anderen Feedbacktypen eingeschränkt herangezogen werden kann. Viele Studien weisen keine ausreichende Studienqualität auf oder sind nicht als Therapiestudie konzipiert worden. Außerdem ist keine im Review eingeschlossene Studie mit dem Ziel eines Feedback-Vergleichs durchgeführt worden, so dass dieser Aspekt nachträglich kaum analysiert werden kann.

Diskussion

Grundlagenwissenschaftliche Befunde weisen darauf hin, dass eine höhere motorische Hirnaktivität durch die Darstellung virtueller Körper möglich ist. Dies könnte für eine effektivere motorische Therapie neurologischer Patientinnen und Patienten genutzt werden.

Die Ergebnisse der systematischen Betrachtung von Publikationen mit VR-Interventionen lassen erkennen, dass in den Forschungsarbeiten zur Implementierung virtueller Darstellungen in der motorischen Neurorehabilitation der Einsatz von Avataren meist unsystematisch erfolgt. Dies bezieht sich auch auf den Einsatz von visuellen Rückkopplungen an Rehabilitationsgeräten. Auch hier wurde bislang der eigenständige Effekt von Avataren zur Unterstützung des motorischen Lernens noch nicht ausreichend untersucht.

Zusammenfassend wird die Art der Visualisierung bisher nicht als gesonderter Faktor berücksichtigt und es bedarf weiterer Forschung um diesen Aspekt im Detail zu untersuchen. Derartige Forschungsergebnisse könnten die Wirkungsweise von VR-Interventionen für zukünftige Anwendungen zielgerichtet optimieren.

Literatur

- Adamovich, S. V., August, K., Merians, A., & Tunik, E. (2009). A virtual reality-based system integrated with fmri to study neural mechanisms of action observation-execution: A proof of concept study. *Restorative neurology and neuroscience*, 27(3), 209–223. doi:10.3233/RNN-2009-0471
- Altschuler, E. L. (2005). Interaction of vision and movement via a mirror. *Perception*, 34(9), 1153–1155.
- Cameirao, M. S., Bermudez i Badia, S., Duarte, E., Frisoli, A., & Verschure, P. F. M. J. (2012). The combined impact of virtual reality neurorehabilitation and Its Interfaces on upper extremity functional recovery in Patients With Chronic Stroke. *Stroke*, 43(10), 2720–2728. doi:10.1161/STROKEAHA.112.653196
- Cameirao, M. S., Bermudez i Badia, S., Duarte Oller, E., & Verschure, P. F. (2010). Neurorehabilitation using the virtual reality based Rehabilitation Gaming System: methodology, design, psychometrics, usability and validation. *Journal of NeuroEngineering and Rehabilitation*, 7(1), 48. doi:10.1186/1743-0003-7-48
- Connelly, L., Jia, Y., Toro, M. L., Stoykov, M. E., Kenyon, R. V., & Kamper, D. G. (2010). A pneumatic glove and immersive virtual reality environment for hand rehabilitative training after stroke. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(5), 551–559. doi:10.1109/TNSRE.2010.2047588
- da Silva Cameirão, M., Bermúdez i Badia, S., Duarte, E., & Verschure, P. F. M. J. (2011). Virtual reality based rehabilitation speeds up functional recovery of the upper extremities after stroke: A randomized controlled pilot study in the

- acute phase of stroke using the Rehabilitation Gaming System. *Restorative Neurology and Neuroscience*, 29(5), 287–298.
- Dohle, C., Stephan, K. M., Valvoda, J. T., Hosseiny, O., Tellmann, L., Kuhlen, T., Seitz, R. J. (2011). Representation of virtual arm movements in precuneus. *Experimental brain research*, 208(4), 543–555. doi:10.1007/s00221-010-2503-0
- Ferreira dos Santos, L., Schmidt, H., Krüger, J., & Dohle, C. (in press). Visualization of virtual reality neurological motor rehabilitation of the upper limb - A systematic review. *IEEE Xplore*.
- Kam, N., Struzik, J., Jarus, T., & Rand, D. (2012). Is the Nintendo Wii suitable for stroke rehabilitation? A pilot feasibility and usability study. *Israel Journal of Occupational Therapy*, 21(1), E3–E23.
- Laver, K. E., George, S., Thomas, S., Deutsch, J. E., & Crotty, M. (2011). Virtual reality for stroke rehabilitation. *Cochrane database of systematic reviews (Online)*, (9), CD008349. doi:10.1002/14651858.CD008349.pub2
- Taylor, M. J. D., McCormick, D., Shawis, T., Impson, R., & Griffin, M. (2011). Activity-promoting gaming systems in exercise and rehabilitation. *Journal of rehabilitation research and development*, 48(10), 1171–1186.
- Thieme, H., Mehrholz, J., Pohl, M., Behrens, J., & Dohle, C. (2012). Mirror therapy for improving motor function after stroke. *Cochrane database of systematic reviews (Online)*, 3, CD008449. doi:10.1002/14651858.CD008449.pub2
- Villiger, M., Chandrasekharan, S., & Welsh, T. N. (2011). Activity of human motor system during action observation is modulated by object presence. *Experimental brain research*, 209(1), 85–93. doi:10.1007/s00221-010-2522-x

Exploring Surprise: Influencing User Experience through Unexpected Events

Alice M. Gross¹

User Experience, Surprise, Emotional Design, User Centered Design

Abstract

When interacting with technical products, users rely on their former experiences with similar products, earlier interaction episodes, and their general knowledge about the product or similar ones. Based on these sources of information, expectations about the interaction are formed. As it has been shown in classical emotional psychology studies, the frustration of expectations leads to surprise. Depending on a person's desires, a surprise can be positive or negative. Surprise as a positive emotion has been shown to be beneficial in classical product design, triggering interest and amusement in users. People feel drawn to surprising and thus interesting product designs, whereas little research has been done about negative surprise in product design. The effects of positive and negative surprise have not been extensively studied in digital interactive products. Whether these effects can be transferred to the user experience design of digital products will be discussed in this paper.

1 Introduction

Interacting with technical products has become an important part of our everyday life. Whether this interaction is a positive and enjoyable experience highly depends on the product's usability, as well as on the resulting user experience (UX).

1.1 The CUE model

Mahlke and Thüring (2007) identified three factors of an interaction with a product which influences the emergence of UX: In their Components of User Experience model (CUE model), they describe how users perceive instrumental and non-instrumental qualities of a product during interaction. During these cognitive processes, they gather knowledge about a product's strengths and weaknesses, its design, feel, and features. These cognitive procedures are accompanied by emotions which may be positive or negative depending on the quality of the interaction. Together, cognition and emotion constitute the users' overall experience that evolves from their actions and the responses of the system (Gross & Thüring, 2013).

Besides these three components of UX, other authors highlight the importance of expectations a user has about the interaction with a product. Pohlmeier (2011) discusses the importance of anticipated experience for UX, whereas Karapanos (2011)

¹ Technische Universität Berlin, Zentrum Mensch-Maschine-Systeme, Graduiertenkolleg prometei

states that even naive users have formed expectations about a products' behavior when in use.

2 The Case of Surprise

Emotions have been the focus of scientific debates for a long time (Kleinginna & Kleinginna, 1981). One special case of emotion is surprise. The reason for this is that contrary to other emotions surprise has one unique feature: depending on a person's specific desires about the outcome of a situation, surprise can be either neutral, accompanied by a positive emotion or accompanied by a negative emotion (Ortony, Clore & Collins, 1990). The neutrality of surprise makes it a controversial topic in the field of psychology since some researchers claim an emotion needs to be either positive or negative. Ortony et al. define surprise without a positive or negative affect as being merely an awareness of unexpectedness and thus not an emotion (1990). Nevertheless, they call surprise accompanied by a positive emotion *pleasant surprise*, and surprise accompanied by a negative emotion *unpleasant surprise*. We will treat surprise as being an emotion of either positive or negative affect.

2.1 The Development of surprise

In its formation, the emotion of surprise is closely linked to the concept of expectation formation. If these expectations are disconfirmed the result is surprise (Reisenzein, 2008). In his paper about the belief-desire theory of emotion, Reisenzein states that a positive surprise is the result of expectation frustration combined with desire fulfillment. Negative surprise in contrast is the result of expectation frustration and desire frustration at the same time (see figure 1). Its unique features make surprise especially interesting for design research.

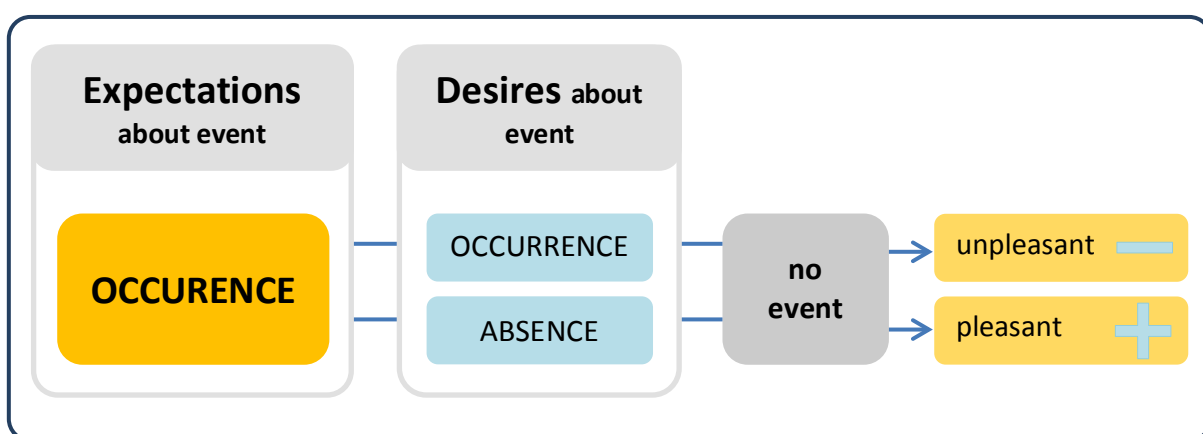


Fig. 1: Schematic development of pleasant/unpleasant surprise (according to Reisenzein, 2008)

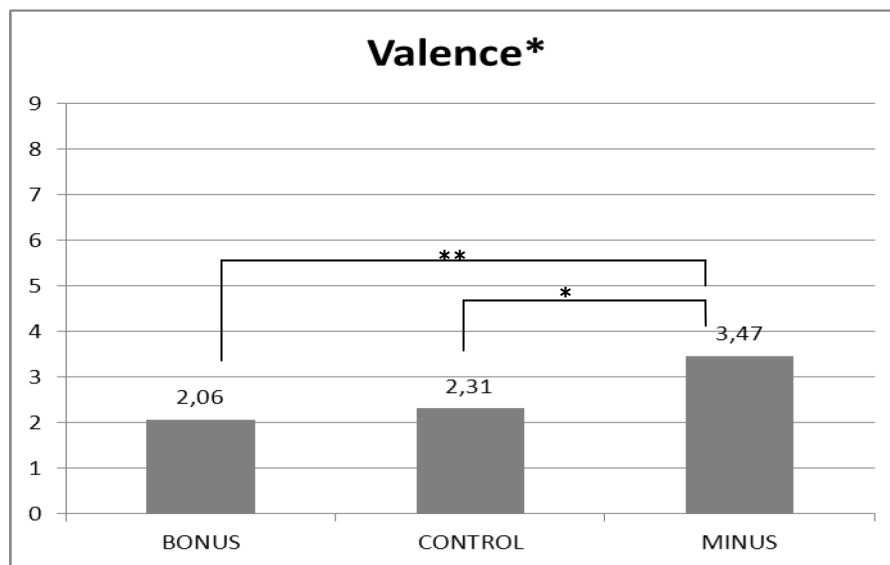
Surprise and Product Design

Research on surprising product design has shown a beneficial effect of surprise for users: Handling surprising products results in an interaction that is more enjoyable, more interesting, and easy to remember (Ludden, Schifferstein & Hekkert, 2009). Product designers have made use of the benefits of pleasant surprise for instance when designing tangible products (Ludden et al., 2009). They were able to demonstrate the beneficial effect of surprise by creating products that had similar visual appearances but differed in their tactual characteristics. By creating these visual-tactual incongruities, they were able to provoke surprise reactions.

While pleasant surprise has been studied extensively in classical product design, not many researchers have actively explored it as a design factor for digital, interactive products. Although some studies refer to surprise related concepts, like WOW, delight or appraisal (Desmet, Porcelijn & van Dijk, 2007; Mori & Inoue, 2004; Väänänen-Vainio-Mattila, Palvianen, Pakarinen, Lagerstam & Kangas, 2011), most research was constrained to non-interactive products.

Experimental findings

To explore a possible transfer of the beneficial effects of positive surprise from classical product design to user experience design for digital interactive products, a study was carried out (for a detailed description see Gross & Thüring, 2013). Besides the beneficial effects of positive surprise, possible harmful effects of negative surprise were investigated, as well. By using the classical computer game Tetris, the impact of surprise on UX ratings was investigated. Participants played different versions of a regular Tetris game, experiencing a sudden loss of point (negative surprise) or a sudden gain of points (positive surprise). A control group played the same game without any surprising events. The success of the surprise manipulation was controlled by comparing reaction times after the appearance of a stone combined with the unexpected event. Positive surprise lead to elevated UX ratings compared to control group ratings whereas negative surprise lead to lower UX ratings. Only the difference between negative surprise and control group were statistically significant (see figure 2 for SAM Valence ratings).



As has been mentioned above, when trying to use surprise as a design element in user experience design, the beneficial effects of positive surprise could not be replicated. What was shown indeed are the profound harmful effects of negative surprise. When interpreting these results, one has to keep in mind that these harmful effects on UX ratings cannot be attributed to bad usability or malfunctioning of the game. All the Tetris games being played during the experiment were functioning identically with the exact same design being implemented. Avoiding negative surprise is absolutely crucial when designing digital products, at least when designing digital games.

Literatur

- Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1), 49-59.
- Desmet P., Porcelijn R., van Dijk M. (2007). Emotional design; Application of a Research-Based Design Approach. *Knowledge, Technology & Policy*, 20(3), 141–155.
- Gross, A., & Thüring, M. (2013). Encountering the Unexpected: Influencing User Experience through Surprise. Shlomo Berkovsky, S., Herder, E., Lops, P. & Santos, O. C. (Eds): *Proc. of UMAP Annual Conference*, Rome, Italy: Springer. 3-9.
- Hassenzahl, M., Burmester, M., Koller, F., (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Ziegler, J., Szwillus, G. (Eds.), *Mensch & Computer 2003. Interaktion in Bewegung*. B.G. Teubner, Stuttgart, Leipzig, pp. 187–196
- Jordan, P.W. (1997). Putting the pleasure into products. *IEE Review* 249-252

- Karapanos, E., Zimmerman, J., Forlizzi, J. & Martens, J.-B. (2009). User Experience Over Time. An Initial Framework. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Boston, US: ACM Press. 729 -738.
- Kleinginna, P. R. J., & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion*, 5(4), 345–379. doi:10.1007/BF00992553
- Ludden, G. D. S., Schifferstein, H. N. J., Hekkert, P. (2009). Visual-tactual incongruities in products as sources of surprise. *Empirical Studies of the Arts*, 27(1), 61-87
- Mahlke, S., Thüring, M. (2007). Studying antecedents of emotional experiences in interactive contexts. In: *Proc. Conference on Human-Computer Interaction*, pp. 915-918. ACM Press, New York (2007)
- Mori, H., Inoue, J. (2004). Jigsaw Panel: A Tangible Approach for Delightful Human-Computer Interaction. In *Proc. of SICE Annual Conference*, Sapporo, Japan. 1579-1582.
- Norman, D. (2004). *Emotional Design: Why we love (or hate) everyday things*. Basic Books, New York
- Ortony, A., Clore, G. L., & Collins, A. (1990). *The Cognitive Structure of Emotions*. Cambridge University Press.
- Väänänen-Vainio-Mattila, K., Palvianen, J., Pakarinen, S., Lagerstam, E., Kangas, E.: User Perception of Wow Experiences and Design Implications for Cloud Services. In: *DPPI '11 Proceedings of the 2011 Conference on Designing Pleasurable Products*, ACM Press, New York (2011)

Asymmetrische Unsicherheitsakzeptanz im Umgang mit Entscheidungsunterstützungssystemen

Torsten Günzler¹ und Dietrich Manzey²

Schlüsselwörter: Entscheidungsunterstützungssystem, Alarm, Entscheidungen unter Unsicherheit, Unsicherheitsakzeptanz, Prüfverhalten

Zusammenfassung

Automatisierte Entscheidungsunterstützung im Kontext komplexer technischer Systeme geht aufgrund technischer Einschränkungen häufig mit Unsicherheiten einher, wie z. B. anhand von Alarmsystemen in Kraftwerken deutlich wird. Nutzer haben daher oft die Möglichkeit Hinweise solcher Entscheidungsunterstützungssysteme zu überprüfen. Bisher ist unklar, wie sich das Prüfverhalten von Nutzern gestaltet und durch was es beeinflusst wird. Jedoch wurde gezeigt, dass ein geringes Maß an Unsicherheit meist ohne Überprüfung akzeptiert wird, wenn der Hinweis sehr wahrscheinlich korrekt ist. Ist er hingegen sehr wahrscheinlich falsch, wird er meist überprüft. Zwei potentiell kausale Einflussfaktoren für dieses asymmetrische Prüfverhalten wurden im Rahmen von zwei experimentellen Studien untersucht. Gemäß den Befunden kann die starke Asymmetrie im Prüfverhalten weder durch eine Asymmetrie in der subjektiven Risikowahrnehmung, noch durch das Vorhandensein eines Alarmkontexts erklärt werden. Weitere Erklärungsansätze werden vorgestellt.

Einleitung

Bei der Steuerung und Überwachung komplexer technischer Systeme werden Menschen häufig durch Automationen unterstützt, die Systemzustände diagnostizieren und kommunizieren. Solche Subsysteme können als Entscheidungsunterstützungssysteme betrachtet werden, da ein effizientes Entscheidungsverhalten häufig erst durch die Hinweise dieser Systeme ermöglicht wird. Sie liefern damit häufig einen wichtigen Beitrag für den Betrieb komplexer Systeme. Besonders deutlich wird dies am Beispiel von Alarmsystemen, ohne die eine sichere und effiziente Steuerung vieler technischer Anlagen durch den Menschen nicht denkbar wäre. Alarmsysteme verdeutlichen jedoch auch eine typische Problematik im Umgang mit automatisierter Entscheidungsunterstützung, da ihre Nutzung fast immer mit Entscheidungen unter Unsicherheit einhergeht. So sind Alarmsysteme durch die technischen Randbedingungen selten absolut zuverlässig und konfrontieren den Nutzer daher mit einem gewissen Grad an Unsicherheit. Diese Unsicherheit kann

¹ Technische Universität Berlin, Zentrum Mensch-Maschine-Systeme, Graduiertenkolleg prometei

² Technische Universität Berlin, Institut für Psychologie und Arbeitswissenschaft, Lehrstuhl für Arbeits-, Ingenieur- und Organisationspsychologie

zu Verzerrungen im Entscheidungsverhalten des Nutzers führen und dadurch seine Steuerungs- und Überwachungsleistung beeinträchtigen (Parasuraman & Riley, 1997). Ein denkbarer Lösungsansatz für diese Problematik ist die Bereitstellung von zusätzlichen Systeminformationen, die auf Wunsch abgefragt werden können um eine Alarmmeldung zu überprüfen (z. B. Rovira, McGarry & Parasuraman, 2007). Es ist bisher jedoch weitgehend unklar, wie Nutzer von Alarmsystemen – und Entscheidungsunterstützungssystemen im Allgemeinen – mit solchen Prüfmöglichkeiten umgehen. Gérard und Manzey (2010) lieferten erste Hinweise dafür, dass das Ausmaß an objektiver Unsicherheit in keinem direkten Zusammenhang zum Prüfverhalten steht. In ihrem Experiment akzeptierten Versuchspersonen ein geringes Maß an Unsicherheit (10%) bei einem mit hoher Wahrscheinlichkeit (90%) richtigen Alarm eher, als bei einem mit hoher Wahrscheinlichkeit (90%) falschen Alarm. Die Möglichkeit, Alarmmeldungen durch das Prüfen von Rohdaten zu verifizieren, wurde von Versuchspersonen wesentlich häufiger genutzt um Entscheidungen im Widerspruch zu einer Alarmmeldung abzusichern, als Entscheidungen im Einklang mit einer Alarmmeldung, obwohl Fehlerwahrscheinlichkeiten und -kosten identisch waren. Es stellt sich somit die Frage, welche Faktoren das Prüfverhalten beeinflussen und wodurch die Asymmetrie im Prüfverhalten hervorgerufen wird.

Asymmetrie als Folge der subjektiven Risikowahrnehmung

Eine Erklärung für das Phänomen der asymmetrischen Prüfquoten könnte eine Asymmetrie des subjektiv wahrgenommen Risikos in den zwei Entscheidungssituationen sein. So bestand die Aufgabe der Versuchsteilnehmer von Gérard und Manzey (2010) in der Steuerung einer „chemischen Anlage“. In der Teilaufgabe, die durch das Alarmsystem unterstützt wurde, mussten Reaktionsbehälter überwacht werden. Diese konnten sich entweder in einem normalen oder einem kritischen Zustand befinden. Das Alarmsystem gab eine Diagnose pro Behälter: entweder eine grüne „OK“-Meldung oder einen roten Alarm. Unabhängig von der Korrektheit dieser Meldungen konnten Versuchspersonen zwei verschiedene Fehler unterlaufen. Sie konnten von einem Normalzustand ausgehen, wenn in Wirklichkeit ein kritischer Zustand vorlag und umgekehrt. Während Teilnehmer für das Verpassen eines kritischen Zustandes aufgrund der Coverstory (Überwachung von Reaktionsbehältern in einer chemischen Anlage) eventuell auf potentiell katastrophale Folgen schlossen (z. B. Explosion, Kontamination), dürfte dies für die fälschliche Annahme eines kritischen Zustandes bei in Wirklichkeit vorliegendem Normalzustand kaum der Fall gewesen sein. Somit könnten Versuchspersonen allein wegen der Coverstory einen Fehler als wesentlich schlimmer bewertet haben als den anderen, auch wenn sich die vorgegebenen Punktabzüge für beide Fehler nicht unterschieden. In einer aktuellen Untersuchung wurde daher eine neutrale Coverstory für die gleiche Aufgabe gewählt („Etikettierungsanlage in einer Brauerei“, für eine detaillierte Beschreibung s. Günzler & Manzey, in Druck). Die beiden möglichen Fehler entsprachen inhaltlich nun einer falschen Etikettierung, was

auch subjektiv einer symmetrischen Risikoverteilung entsprechen sollte. Es zeigte sich bei dieser neutralen Coverstory jedoch das identische Verhaltensmuster wie bei Gérard und Manzey (2012). Während Hinweise, die zu 90% korrekt waren, kaum überprüft wurden, wurden Hinweise, die zu 90% falsch waren, sehr häufig überprüft, obwohl in beiden Fällen die gleiche Unsicherheit von 10% vorlag und Kosten für beide Arten von Fehlentscheidung identisch waren (s. Abb. 1). Somit lässt sich die Asymmetrie im Prüfverhalten nicht mit der subjektiven Risikostruktur erklären.

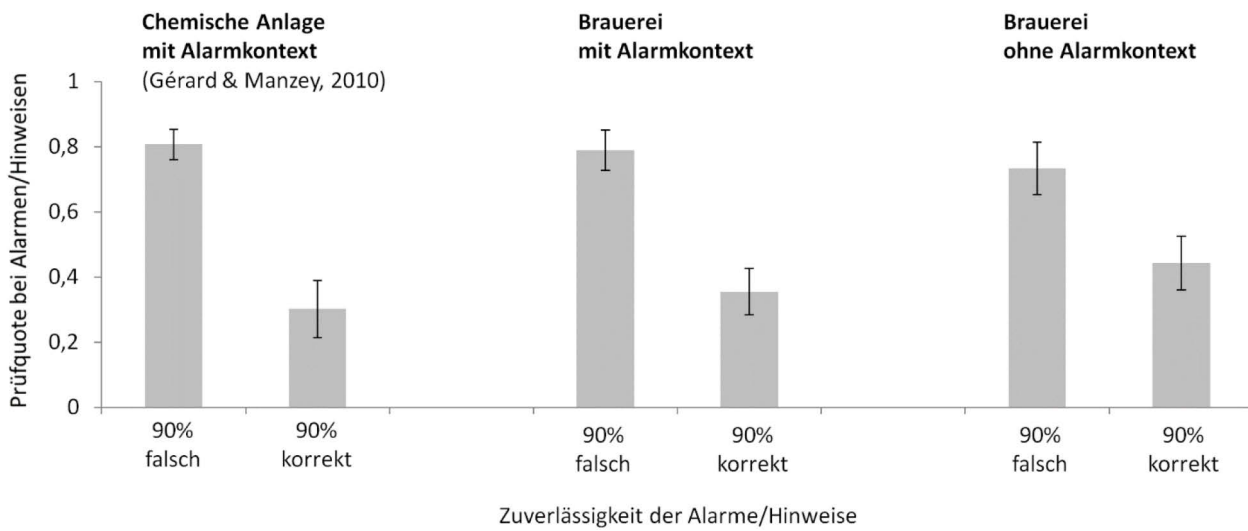


Abb.1: Prüfquoten bei Alarmen bzw. analogen neutralen Hinweisen in Abhängigkeit von der Zuverlässigkeit der Hinweise sowie der Art der Coverstory und dem Vorhandensein eines Alarmkontextes (Fehlerbalken = 1 Standardfehler)

Asymmetrie als Folge des Alarmkontexts

Eine andere denkbare Erklärung für das Phänomen der asymmetrischen Akzeptanz von Unsicherheit könnte der Alarmkontext sein. So zeigen Alarme erfahrungsgemäß wichtige Ereignisse an, deren Verpassen potentiell schlimme Folgen haben kann. Demnach könnte bereits die Verknüpfung eines zu detektierendem Ereignisses mit einem Alarm dazu führen, dass dieses Ereignis als vergleichsweise gefährlich wahrgenommen wird. In diesem Falle wäre die Asymmetrie der Unsicherheitstoleranz auf den Alarmkontext begrenzt und würden nicht bei neutralen Systemmeldungen, die nicht mit Alarmen assoziiert sind, auftreten.

Diese Annahme wurde im Rahmen einer weiteren Untersuchung überprüft. Versuchspersonen bearbeiteten die gleiche Aufgabe wie zuvor beschrieben. Allerdings wurde ihnen nun in einer neuen Bedingung das Alarmsystem als Assistenzsystem vorgestellt. Bisherige „Alarme“ wurden neutral als „Systemhinweise“ beschrieben und auch die Farbgebung der Hinweise wurde neutral gestaltet (hellblau/dunkelblau statt grün/rot), um Assoziationen mit Alarmen zu vermeiden.

Interessanterweise bestand selbst in dieser neutralen Bedingung das asymmetrische Prüfverhalten fort (s. Abb. 1). Auch wenn das Ausmaß der Asymmetrie auf

deskriptiver Ebene leicht abgeschwächt erscheint, unterscheidet es sich nicht signifikant von denen der vorherigen Studien. Die starke Asymmetrie in den Prüfquoten lässt sich somit auch nicht durch den Alarmkontext erklären.

Fazit und Ausblick

Im Rahmen zweier experimenteller Studien konnten zwei denkbare Ursachen für den Effekt der asymmetrischen Unsicherheitstoleranz im Umgang mit automatisierter Entscheidungsunterstützung ausgeschlossen werden. Asymmetrisches Prüfverhalten wird offenbar weder durch eine subjektive Asymmetrie der Kosten von Entscheidungsfehlern, noch durch die Einbindung von Automationshinweisen in einen Alarmkontext verursacht. Damit ergibt sich weiterer Forschungsbedarf.

Im Rahmen zukünftiger Untersuchungen könnte z. B. der Einfluss von Defaults untersucht werden, welche in solchen Entscheidungssituationen, die auf der Analyse von Systemzuständen beruhen, häufig auftreten. So können sowohl Defaults inhaltlicher Natur existieren (wenn z. B. einer von mehreren möglichen Zuständen der „Normalzustand“ ist) als auch systemimmanent sein (wenn das System standardmäßig von einem bestimmten Zustand ausgeht). Bisherige Befunde zum asymmetrischen Prüfverhalten basieren auf einem System, welches mit systemimmanenten Default arbeitete, sodass sich die Frage stellt, ob bereits allein die diesbezügliche Systemgestaltung für das asymmetrische Prüfverhalten verantwortlich zeichnet.

Ferner könnte Personen nach dem Prinzip der geteilten Verantwortung eine Entscheidung im Einklang mit einem Systemhinweis einfacher fallen als im Widerspruch zu einem Systemhinweis. Im Falle einer Fehlentscheidung könnte die empfundene Verantwortung für den Fehler leichter wiegen, wenn die Entscheidung „gemeinsam“ mit dem System anstatt entgegen der Systemempfehlung gefällt wurde. Demnach könnte eine erhöhte Prüfquote bei Entscheidungen entgegen einem Hinweis auf eine subjektiv als höher empfundene Verantwortung zurückzuführen sein. Der höheren Verantwortung würde also Rechnung getragen werden, indem man die Entscheidung durch Prüfen absichert.

Auch über das Phänomen der Asymmetrie hinaus stellt sich die Frage, inwiefern die Überprüfung von Hinweisen von Entscheidungsunterstützungssystemen durch den Nutzer generellen Verzerrungen unterliegt. Nicht zuletzt in Hinblick auf eine sinnvolle Gestaltung solcher Systeme gilt es ein besseres Verständnis dieser Entscheidungsprozesse anzustreben.

Literatur

Gérard, N. & Manzey, D. (2010) Are false alarms not as bad as supposed after all? A study investigating operators' responses to imperfect alarms. In D. de Waard, A. Axelsson, M. Berglund, B. Peters & C. Weikert (Hrsg.), *Human*

Factors. A system view of human, technology and organisation (S. 55-70). Maastricht: Shaker Publishing.

Günzler, T. & Manzey, D. (in Druck). Asymmetries in human tolerance of uncertainty in interaction with alarm systems: Effects of risk perception or evidence for a general commission bias? *Proceedings of the HFES 57th Annual Meeting 2013*. Santa Monica: Human Factors and Ergonomics Society.

Parasuraman, R. & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(2), 230-253.

Rovira, E., McGarry, K. & Parasuraman, R. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49, 76-87.

Adaptive Regelungsalgorithmen für das robotergestützte motorische Training der oberen Extremitäten für Patienten nach Schlaganfall: Literatur-Review

Ekaterina Ivanova¹, Henning Schmidt², Jörg Krüger³

Schlüsselwörter: Roboter, Adaptive Regelung, Mensch-Maschine-Interaktion, Rehabilitation, Schlaganfall

Zusammenfassung

Im Bereich der robotergestützten Neurorehabilitation bestimmen die in der Robotersteuerung implementierten Regelungsalgorithmen, wie die Interaktion zwischen Patient und dem Übungsgerät erfolgt, was ein wichtiger Faktor für Rehabilitationserfolg sein kann. Adaptive „Assist-as-needed“-Ansätze berücksichtigen die motorischen Fähigkeiten vom Patienten und gewährleisten nur so viel Unterstützung, wie er braucht. Somit wird dem Patienten die Möglichkeit gegeben, an der Therapie aktiver teilzunehmen und, infolgedessen, größere Fortschritte im Wiedererlernen willkürlicher Bewegungen zu machen. Um die existierenden Lösungen zu sichten und zu systematisieren, wurde eine Literatursuche in den Datenbanken IEEE Xplore, Medline und PubMed durchgeführt. Es wurden 1877 relevante Publikationen ermittelt. Nach der Analyse von Metadaten und Volltexten wurden davon 18 Publikationen in das Review eingeschlossen. Nach der Analyse von in den Veröffentlichungen dargestellten Regelungsalgorithmen wurden die Ergebnisse nach dem Typ der Regelung und der Methode, die den Grad der Unterstützung vom Roboter einstellt, klassifiziert.

Einführung

Die robotergestützte Bewegungstherapie für Patienten nach Schlaganfall gewinnt immer mehr an Bedeutung. Die Effektivität der Rehabilitationsroboter, die vor allem die Unterstützung von Patienten bei der Wiederholung der Bewegungen übernehmen, ist in mehreren Studien bestätigt (Prange et al., 2006, Kwakkel et al., 2008, Hesse et al., 2004) und ist nach manchen Ergebnissen höher als bei konventionellen Ansätzen (Lum, et al., 2002). Die Wirksamkeit der Reharobotik kann unter anderem durch höhere Genauigkeit und Intensität der ausgeführten Bewegungen begründet werden (Kwakkel et al., 2008), welche die Minderung der motorischen Beeinträchtigung steigern (Kwakkel et al., 1999, Han et al., 2013).

Die Übungen mit Rehabilitationsgeräten basieren in der Regel darauf, dass die betroffene Hand vom Roboter entlang einer Übungstrajektorie passiv oder mit be-

¹ Graduiertenkolleg prometei, Arbeitsgruppe Rehabilitationsrobotik (IPK/TU Berlin), Technische Universität Berlin

² Arbeitsgruppe Rehabilitationsrobotik (IPK/TU Berlin), Bereich Automatisierungstechnik, Fraunhofer IPK, Berlin

³ Fachgebiet Industrielle Automatisierungstechnik, Technische Universität Berlin

stimmtem Grad an Unterstützung geführt wird (Lum, et al., 2002). Die in den Übungsgeräten implementierten Regelungsalgorithmen dienen dazu, die Konzepte der Neurorehabilitation und die Theorien des motorischen Lernens optimal umzusetzen und zu bestimmen, wie die Interaktion zwischen den Patienten und dem Gerät erfolgt.

Die existierenden Lösungen lassen sich grob in vier Kategorien unterteilen (Marchal-Crespo et al., 2009): i) assistierende Ansätze, ii) challenge-basierte Ansätze, iii) haptische Simulationen der Alltagsaufgaben und iv) Hands-off Coaching. Im Review ist die assistierende Strategie in den Fokus genommen, bei deren es sich um haptische Führung (guidance) handelt. Die haptische Führung ist eine motorische Lernstrategie, die nicht korrekten Bewegungen beim motorischen Lernen durch eine direkte physisch-haptische Unterstützung vorbeugt oder Bewegungsfehler minimiert (Schmidt, Lee, 2005). Im Fall der Schlaganfallrehabilitation sollte die Unterstützung nur so groß wie nötig sein und so schnell wie möglich reduziert werden (O'Sullivan, 2007). Im umgekehrten Fall wird der Mensch die unterstützenden Kräfte in seinen motorischen Plan einbinden und selber den beim Lernen benötigten Aufwand minimieren, was negative Auswirkung auf die Therapiefortschritte haben kann (Marchal-Crespo et al., 2009). Hingegen haben adaptive Strategien das Potential, den Patienten zur aktiven Teilnahme in der Übung zu veranlassen (Riener et al., 2005). Deswegen werden zunehmend adaptive Regelungen in die Rehabilitationsgerätesteuerungen integriert, die die Fähigkeiten der Patienten berücksichtigen, passende Unterstützung gewährleisten und tendenziell bessere Therapieresultate im Vergleich zu passiver Armführung zeigen (Reinkensmeyer et al., 2009).

Es gibt eine große Bandbreite adaptiv assistierender Regelungsalgorithmen im Bereich der Neurorehabilitation. Oft sind sie in den Rehabilitationsroboter für untere Extremitäten implementiert, wie zum Beispiel bei Riener et al., (2005), Emken et al., (2005) oder Hussein et al., (2013). Die Algorithmen basieren in der Regel auf einer fixierten Referenztrajektorie und passen den Unterstützungsgrad individuell an die Patientenfähigkeiten an (Riener et al., 2005). Da die Bewegungsmuster beim Gehen im Gegensatz zu Armbewegungen ähnliche Trajektorien haben, die mittels Ganganalyse ermittelt werden können, sind solche Regelungsansätze besser anwendbar. Die Alltagsbewegungen der Arme sind ziemlich vielfältig, deswegen sollten zusätzlich zur Unterstützungsanpassung die Bewegungstrajektorien ausgewählt bzw. gestaltet werden.

Das Ziel dieses Reviews ist auf Grundlage der Literaturrecherche einen Überblick über den Stand der Technik der in der robotergestützten Rehabilitation verwendeten adaptiven Regelungsalgorithmen zu schaffen und die existierenden Lösungen zu systematisieren. Diese Arbeit stellt nur die Lösungen für obere Extremitäten dar.

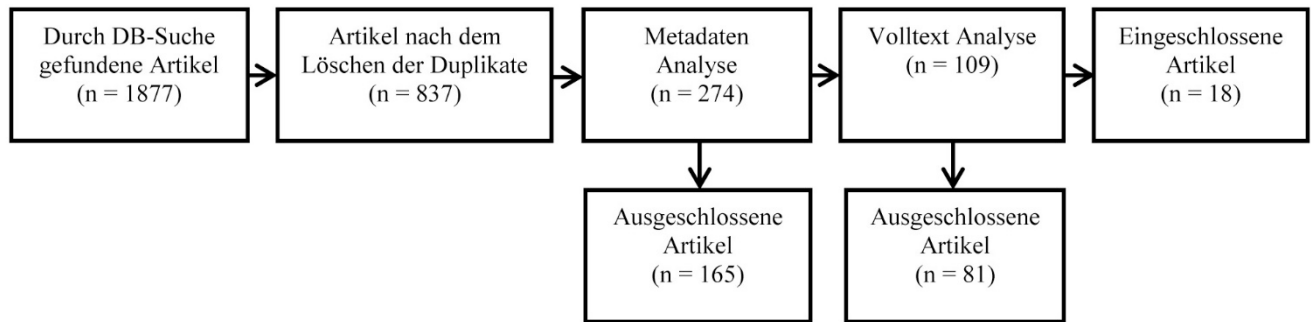


Abb.1: Suche und Screening Prozess

Methodik

Die Literatursuche wurde in den Datenbanken IEEE Xplore, Medline und PubMed durchgeführt. Die Suche erfolgte bis Mai 2013. Bei der Suche wurden folgende Schlüsselwörter verwendet: adaptive, assist-as-needed, control, strategy, algorithm, robot, rehabilitation, stroke, upper extremity, arm. Dabei wurden auch alle Wortkombinationen, die mit den Begriffen „adaptive“, „control“ oder „assist-as-needed“ anfangen, berücksichtigt. Es wurden nur Veröffentlichungen in Englischer oder Deutscher Sprache in Betracht gezogen.

Artikel mit folgenden Kriterien wurden eingeschlossen:

1. es handelt sich um ein roboterbasiertes assistierendes Gerät für die oberen Extremitäten für die Neurorehabilitation,
2. der in der Robotersteuerung angewendete Regelungsalgorithmus ist adaptiv und ist im Artikel mathematisch und inhaltlich beschrieben.

Nur diejenigen Veröffentlichungen, die beide Kriterien erfüllen, wurden berücksichtigt. Publikationen, die zur Steuerung der Roboter externe Messsignale wie z.B. EEG- (Elektroenzephalografie), EMG- (Elektromyografie) oder zur Muskelaktivierung auf FES (Functional electrical stimulation) basierende Systeme beschreiben, wurden aus einer weiteren Betrachtung ausgeschlossen. Außerdem wurden keine Regelungsstrategien berücksichtigt, die als „compliant“ definiert werden können, sowie „softe“ Roboter und Roboter, die die Reaktion auf die Initiierung der Muskelbewegungen vom Patienten gewährleisten (Riener et al., 2005).

Insgesamt wurden 1877 grundsätzlich in Frage kommende Publikationen identifiziert. Nach dem Aussondern von Duplikaten und einer ersten Filterung nach den Titeln blieben 837 Artikel übrig. Zwei Reviewer lasen unabhängig voneinander Titel und die Zusammenfassung der Veröffentlichungen und entschieden, ob jeweilige Publikation relevant sei. Nach der Analyse von Metadaten wurden 109 Artikel ausgewählt und nach der Analyse von Volltexten wurden 80 Publikationen ausgeschlossen. Der Such- und Auswahlprozess ist in Abbildung 1 schematisch dargestellt.

Ergebnisse

Insgesamt wurden 18 Publikationen aus 837 gefundenen in das Review eingeschlossen. Alle Artikel erfüllen beiden Einschlusskriterien. Eine kurze Zusammenfassung der Treffer ist in Tabelle 1 dargestellt.

Tab. 1: Übersicht der eingeschlossenen Publikationen

ID der Publikation	Jahr	Roboter / Konstruktions-typ	Rege-lung	Methode
Balasubramanian et al.	2008	RUPERT / Exoskelett	Position	Iterativ Lernende Regelung, Fuzzylogik
Balasubramanian et al.	2010	RUPERT / Exoskelett	Impe-danz	„Minimal Jerk“ Trajektorie
Erol et al.	2005	PUMA 560 / Endeffektor	Kraft	Künstliches neuronales Netz
Frisoli et al.	2011	PERCRO L-Exos / Exoskelett	Impe-danz	Sliding-Model, Gain Scheduling
Guerrero et al.	2009	Staubli RX90 / Endeffektor	Impe-danz	Bahnsteuerung, „Minimal Jerk“ Trajektorie
Guidali et al.	2011	ARMin III / Exoskelett	Impe-danz	Bahnsteuerung, „Minimal angular Jerk“ Trajektorie
Guidali, Schlink et al.	2011	ARMin III / Exoskelett	Impe-danz	Bahnsteuerung, „Minimal angular Jerk“ Trajektorie, Unterstützungsmodell (Wolbrecht et al.)
Kang et al.	2012	/ Endeffektor	Admittanz	Bahnsteuerung
Mihelja et al.	2007	ARMin / Exoskelett	Impe-danz	„Minimal Jerk“ Trajektorie, Bahnsteuerung
Morbi et al.	2013	/ Exoskelett	Admittanz	„Assistance regulation controller“ (ARC)
Rinderknecht et al.	2011	/ Endeffektor	Impe-danz	Adaptive Oszillator
Ronsse et al.	2011	NEUROExos /	Impe-	Adaptive Oszillator

	0	Exoskelett	danz	
Rosati et al.	2009	Sofia-3, Sofia-4 / Endeffektor	Impe- danz	Bahnsteuerung, Unterstützungsmodell (Wolbrecht et al.)
Trlep et al.	2011	HapticMaster / Endeffektor	Admittan- z	„Minimal Jerk“ Trajek- torie, Bahnsteuerung
Wolbrecht et al.	2007	Pneu-WREX / Exoskelett	Impe- danz	Unterstützungsmodell (Wolbrecht et al.)
Wolbrecht et al.	2008	Pneu-WREX / Exoskelett	Impe- danz	Unterstützungsmodell (Wolbrecht et al.)
Xu et al.	2008	WAM Arm	Impe- danz	Fuzzylogik
Xu et al.	2009	WAM Arm	Impe- danz	Fuzzylogik, neuronales Netz

Die gefundenen Lösungen für die Regelungsgestaltung können nach den verwendeten Eingangs- und Ausgangswerten in vier Hauptkategorien unterteilt werden: Positions-, Kraft- und als Kombination von den beiden Impedanz- und Admittanzregelung.

Eingangs- und Ausgangsgröße der Positionsregelung ist eine Position oder Orientierung des Nutzers bzw. des Endeffektors des Roboters im vordefinierten Koordinatensystem (Mason, 1981). Die Regelung lässt eine Vielzahl von Bewegungsmustern ausführen (Colombo et al., 2005), die Ausgangsbewegungen können dabei als „steif“ bezeichnet werden (Riener et al., 2005). Kraftregelung kann als eine Analogie zur Positionsregelung definiert werden, regelt aber Kraft oder Drehmoment am Endeffektor (Mason, 1981). Die Position- und Kraftregelung sind nur bei jeweils einer Publikation geschrieben worden.

Die Lösungen, die auf Impedanzregelung basieren (13 Publikationen), wurden von Hogan (1985) für Roboteranwendungen vorgeschlagen, die mit der Umgebung bzw. einem Mensch interagieren. Dabei werden Positionsmessdaten (flow) bei der Interaktion von einem Mensch aufgenommen und Force Feedback (effort) wird vom Gerät zurückgegeben. Die alternative Methode zur Impedanz- ist die Admittanzregelung (3 Publikationen). Diese definiert auch dynamische Beziehungen zwischen Position und Kraft, aber nutzt Positionswerte als Input.

Eine weitere Klassifikationsmöglichkeit bezieht sich auf die verwendete Methode, die die Abhängigkeit zwischen Eingangs- und Ausgangsgrößen bestimmt und aus Bewegungsfehlern ein Feedback für Patienten berechnen lässt. In 7 Veröffentlichungen wurde die Bahnsteuerung (path control) mit Kraftfeldern verwendet. Dabei wird eine Referenztrajektorie, die die Soll-Trajektorie von Patient darstellt,

definiert. Um dieser Trajektorie wird ein virtueller Tunnel mit Kraftfeldern geschaffen. Nur wenn ein Patient bei der Bewegungsausführung von der Trajektorie aus dem zulässigen Bereich abweicht, wird der Arm vom Roboter unterstützt (Guidali et al., 2011). Für das Design einer Referenztrajektorie für die Bahnsteuerung wird in 5 Publikationen die von Flash und Hogan (1985) entwickelte Model verwendet, das „minimal jerk trajectory“ genannt wird. Das Model besagt, dass menschliche Bewegungen dazu tendieren, den Ruck (jerk) im Verlauf der Bewegung zu minimieren. Unter dem Ruck wird die Rate der Änderung der Beschleunigung über die Zeit verstanden, was auch die Ableitung der Beschleunigung oder die dritte Ableitung der Position ist. Nach der Theorie wird eine Bewegung glatter, wenn der Parameter J in (1) minimiert wird:

$$J(x) = \int_{t_0}^{t_f} \left(\frac{d^3x}{dt^3} \right)^2 dt \quad (1)$$

Die Unterstützungsanpassung wird in vier gefundenen Publikationen durch eine modelbasierte Methode von Wolbrecht et al. (2008) bestimmt. Diese parametrisiert die Unterstützung von Armbewegungen mit radialer Basisfunktion. Der adaptive Kontroller lernt in Echtzeit ein dynamisches Modell des Armes sowie ein Modell über die Patientenfähigkeiten. Dabei ermöglicht die Verwendung einer Modellbasierten Regelung, Impedanz und Unterstützungskraft separat zu steuern, so dass das Gerät gleichzeitig „compliant“ ist und dem Patienten genug Unterstützung geben kann, um Bewegungen abzuschließen. Die Methode gehört zur Methoden der Iterativ Lernenden Regelung (ILR), die bei wiederholten Bewegungen die Steuerung fehlerbasiert optimiert. Eine alternative ILR-Methode wurde von Balasubramanian et al. (2008) vorgeschlagen. Die Parametrisierung beim System erfolgt über Fuzzylogik-Regeln.

Die auf Fuzzylogik basierende Regelung wird insgesamt in 3 Veröffentlichungen genannt. Die Idee von Fuzzylogik ist die Determinierung linguistisch beschriebener Regeln, die auch nicht eindeutige Informationen und Daten bearbeiten können. (Mamdani, 1974). Die Methode kann auch in Verbindung mit Neuronalen Netzen benutzt werden, die die Fähigkeit zum künstlichen Lernen und zur Abbildung nichtlinearer Modelle haben (Xu et al., 2009). Das künstliche Neuronale Netz, das mit Parametern menschlicher Arme trainiert wurde und die Unterstützungskraft einstellt, ist in der Lösung von Erol (2005) angewendet.

Das Konzept des adaptiven Oszillators wurde in zwei Publikationen beschrieben. Der adaptive Oszillator kann mit externen Signalen synchronisiert werden und deren Parameter werden angepasst, um die Frequenz eines periodischen Eingangssignals zu lernen (Righetti et al., 2009). Die Eigenschaften werden benutzt, um das Verfahren mit dem durch den Benutzer erzeugten Drehmoment bei der Interaktion mit einem Gerät zu synchronisieren (Ronsse et al., 2010).

Zusammenfassung und Diskussion

Das Literatur-Review gibt eine Zusammenfassung von existierenden Lösungen für adaptive Regelungsstrategien im Bereich der robotergestützten Rehabilitation. Die gefundenen Veröffentlichungen wurden nach dem Typ der Regler in vier Kategorien unterteilt: Kraft- (n=1), Positions- (n=1), Impedanz- (n=13), und Admittanzregelung (n=3). Die meisten Systeme basieren auf Impedanzregelung, die eine passende Unterstützung für den Patienten aus dem Positionsfehler bietet. Die Auswahl der Unterstützung erfolgt durch folgende Methoden: Bahnsteuerung mit Kraftfeldern (n=7), iterativ lernende Regelung (n=5), Fuzzylogik (n=3), neuronale Netze (n=2) und adaptiver Oszillator (n=2).

Die Publikationen wurden nur aus einer kleinen Anzahl von Datenbanken erhoben, weshalb die dargestellten Lösungen noch nicht als vollständig zu betrachten sind. Einige Ergebnisse wurden ausgeschlossen, weil Algorithmen nicht vollständig beschrieben wurden und keine ausführliche Information über die zugrunde liegenden Methoden geben. Die Compliance-Regelung und die Systeme, die EMG-Signale nutzen und die im Zusammenhang mit FES funktionieren, sowie die Lösungen für untere Extremitäten wurden im Review nicht berücksichtigt. Als Fortsetzung der Arbeit können weitere Publikationen recherchiert und so die Liste der Treffer erweitert werde.

Literatur

- Balasubramanian, S., Wei, R., He, J. (2008). RUPERT closed loop control design. *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, 3467-3470, doi: 10.1109/IEMBS.2008.4649952.
- Balasubramanian, S., Zhang, H., Buchanan, S., Austin, H., Herman, R., He, J. (2010). Cooperative and active assistance based interactive therapy. *Complex Medical Engineering (CME), 2010 IEEE/ICME International Conference on*, 311-315, doi: 10.1109/ICCME.2010.5558824.
- Emken, J.L. and Bobrow, J.E. and Reinkensmeyer, D.J. (2005). Robotic movement training as an optimization problem: designing a controller that assists only as needed. *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005*, 307-312, doi:10.1109/ICORR.2005.1501108
- Erol, D., Mallapragada, V., Sarkar, N. (2005). Adaptable force control in robotic rehabilitation. *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, 649-654, doi: 10.1109/ROMAN.2005.1513853.
- Flash, T., & Hogan, N. (1985). The coordination of arm movements: an experimentally confirmed mathematical model. *The journal of Neuroscience*, 5(7), 1688-1703.

- Frisoli, A., Sotgiu, E., Procopio, C., Bergamasco, M., Rossi, B., Chisari, C. (2011). Design and implementation of a training strategy in chronic stroke with an arm robotic exoskeleton. *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on*, doi: 10.1109/ICORR.2011.5975512.
- Guerrero, C.D.R., Fraile Marinero, J.C., Turiel, J.P (2009). Robot adaptive behavior to suit patient needs and enable more intensive rehabilitation tasks. *Mechatronics, 2009. ICM 2009. IEEE International Conference on*, doi:10.1109/ICMECH.2009.4957230.
- Guidali, M., Duschau-Wicke, A., Broggi, S., Klamroth-Marganska, V., Nef, T., Riener, R. (2011). A robotic system to train activities of daily living in a virtual environment. *Medical & Biological Engineering & Computing*, 49(10), 1213-1223, doi:10.1007/s11517-011-0809-0.
- Guidali, M., Schlink, P., Duschau Wicke, A., Riener, R. (2011). Online learning and adaptation of patient support during ADL training. *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on*, doi:10.1109/ICORR.2011.5975434.
- Han, C., Wang, Q., Meng, P., Qi, M. (2013). Effects of intensity of arm training on hemiplegic upper extremity motor recovery in stroke patients: a randomized controlled trial. *Clinical Rehabilitation*, 27(1), 75-81.
- Hesse, S., Werner, C., Bardeleben, A. (2004). Der schwer betroffene Arm ohne distale Willküraktivität - ein "Sorgenkind" der Rehabilitation nach Schlaganfall?! *Neurologie and Rehabilitation*, 7, 120-126.
- Hogan, N. (1985). Impedance control: An approach to manipulation. *Journal of Dynamic Systems, Measurement, and Control*, 107, 1-23.
- Hussein, S., Xie, S.Q., Jamwal, P.K. (2013). Adaptive Impedance Control of a Robotic Orthosis for Gait Rehabilitation. *Cybernetics, IEEE Transactions on*, 43, 1025-1034, doi: 10.1109/TSMCB.2012.2222374
- Kang, Y., Jeon, D. (2012). Rehabilitation robot control using the VSD method. *System Integration (SII), 2012 IEEE/SICE International Symposium on*, 192-197, doi: 10.1109/SII.2012.6427313.
- Kwakkel, G., Kollen, B. J., Krebs, H. I. (2008). Effects of Robot-assisted therapy on upper limb recovery after stroke: A Systematic Review. *Neurorehabilitation and Neural Repair*, 22(2), 111-121.
- Kwakkel, G., Wagenaar, R.C., Twisk, J.W.R., Lankhorst, G.J., Koetsier, J.C. (1999). Intensity of leg and arm training after primary middle-cerebral-artery stroke: a randomised trial. *The Lancet*, 354, 191-196.
- Lum, P., Burgar, C., Shor, P., Majmundar, M., Van der Loos, M. (2002). Robot-assisted movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke. *Archives of Physical Medicine and Rehabilitation*, 83, 952-959.

- Lum, P., Reinkensmeyer, D., Mahoney R., Rymer W.Z., Burgar C. (2002). Robotic Devices for Movement Therapy After Stroke: Current Status and Challenges to Clinical Acceptance. *Topics in Stroke Rehabilitation*, 8(4), 40-53.
- Mamdani, E. H. (1974) "Application of fuzzy algorithms for control of simple dynamic plant", *Proceeding of the Institution of Electrical Engineers*, 121(12), 1588-1588.
- Marchal-Crespo, L., Reinkensmeyer, D. J. (2009). Review of control strategies for robotic movement training after neurologic injury. *Journal of NeuroEngineering and Rehabilitation*, 6:20, doi:10.1186/1743-0003-6-20
- Mason, M. (1981). Compliance and force control for computer controlled manipulators. *Systems, Man and Cybernetics, IEEE Transactions on*, 11(6), 418-432.
- Mihelja, M., Nefb, T., Riener, R. (2007). A novel paradigm for patient-cooperative control of upper-limb rehabilitation robots. *Advanced Robotics*, 21(8), 843-867, doi: 10.1163/156855307780851975.
- Morbi, A., Ahmadi, M., Chan, A.D.C., Langlois, R. (2013). Stability-Guaranteed Assist-as-Needed Controller for Powered Orthoses. *Control Systems Technology, IEEE Transactions on*, doi:10.1109/TCST.2013.2259593.
- O'Sullivan, S. B. & Schmitz, T. J. (2007). *Physical Rehabilitation*. Vol. 5: F.A. Davis Company.
- Prange, G. B., Jannink, M. J. A., Groothuis-Oudshoorn, C. G. M., Hermens, H. J., Ijzerman, M. J. (2006). Systematic review of the effect of robot-aided therapy on recovery of the hemiparetic arm after stroke. *Journal of Rehabilitation Research and Development*, 43(2), 171-184.
- Reinkensmeyer, D. J., Patton, J. L. (2009). Can Robots Help the Learning of Skilled Actions? *Exercise and Sport Sciences Reviews*, 37(1), 43-51.
- Righetti, L., Buchli, J., Ijspeert, A. J. (2009). Adaptive frequency oscillators and applications. *Open Cybernetics & Systemics Journal*, 3, 64-69.
- Riener, R., Lünenburger, L., Jezernik, S., Anderschitz, M., Colombo, G., Dietz, V. (2005). Patient-cooperative strategies for robot-aided treadmill training: first experimental results. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13, 380-394, doi:10.1109/TNSRE.2005.848628.
- Rinderknecht, M.D., Delaloye, F.A., Crespi, A., Ronsse, R., Ijspeert, A.J. (2011). Assistance using adaptive oscillators: Robustness to errors in the identification of the limb parameters. *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on*, 1-6, doi:10.1109/ICORR.2011.5975351.
- Ronsse, R., Vitiello, N., Lenzi, T., van den Kieboom, J., Carrozza, M.C, Ijspeert, A.J (2010). Adaptive oscillators with human-in-the-loop: Proof of concept for assistance and rehabilitation. *Biomedical Robotics and Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS International Conference on*, 668-674, doi: 10.1109/BIOROB.2010.5628021.

- Rosati, G., Zanotto, D., Secoli, R., Rossi, A. (2009). Design and control of two planar cable-driven robots for upper-limb neurorehabilitation. *Rehabilitation Robotics, 2009. ICORR 2009. IEEE International Conference on*, 560-565, doi: 10.1109/ICORR.2009.5209551.
- Schmidt, R.A. & Lee, T.D. (2005). *Motor Control and Learning: A Behavioral Emphasis*. Vol. 4: Human Kinetics.
- Trlep, M., Mihelj, M., Munih, M., Puh, U. (2011). Rehabilitation robot for unimanual and bimanual training of hemiparetic subjects. *Virtual Rehabilitation (ICVR), 2011 International Conference on*, 1-6, doi: 10.1109/ICVR.2011.5971861.
- Wolbrecht, E.T., Chan, V., Le, V., Cramer, S.C., Reinkensmeyer, D.J., Bobrow, J.E. (2007). Real-time computer modeling of weakness following stroke optimizes robotic assistance for movement therapy. *Neural Engineering, 2007. CNE '07. 3rd International IEEE/EMBS Conference on*, 152-158, doi: 10.1109/CNE.2007.369635.
- Wolbrecht, E.T., Chan, V., Reinkensmeyer, D.J., Bobrow, J.E. (2008). Optimizing Compliant, Model-Based Robotic Assistance to Promote Neurorehabilitation. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 16, 286-297, doi: 10.1109/TNSRE.2008.918389.
- Xu, G., Song, A. (2008). Fuzzy Variable Impedance Control for Upper-Limb Rehabilitation Robot. *Fuzzy Systems and Knowledge Discovery, 2008. FSKD '08. Fifth International Conference on*, 3, 216-220, doi: 10.1109/FSKD.2008.101.
- Xu, G., Song, A. (2009). Adaptive impedance control based on dynamic recurrent fuzzy neural network for upper-limb rehabilitation robot. *Control and Automation, 2009. ICCA 2009. IEEE International Conference on*, 1376-1381, doi: 10.1109/ICCA.2009.5410286

Einfluss von Alarmen chirurgischer Assistenzsysteme auf Leistung, Beanspruchung und Situationsbewusstsein von Chirurgen

Maria Luz¹ und Dietrich Manzey²

Schlüsselwörter: chirurgische Navigationssysteme, Automation, Alarme, Assistenzsysteme

Zusammenfassung

Chirurgische Navigationssysteme erleichtern Chirurgen die räumliche Orientierung in der Patientenanatomie und unterstützen ihn bei der Identifizierung von anatomischen Strukturen. Der neuere Ansatz, das sogenannte Navigated Control (NC)-System bietet eine zusätzliche Sicherheitsfunktion, indem es die chirurgische Fräse automatisch ausschaltet, wenn diese sensiblen Strukturen zu nah kommt. Dadurch wird die unbeabsichtigte Schädigung der sensiblen Strukturen aktiv verhindert. Zusätzlich warnt ein optischer Alarm auf dem Navigationsbildschirm vor möglichen Verletzungsrisiken. In zwei experimentellen Simulationsstudien mit erfahrenen und unerfahrenen Chirurgen wurden die Auswirkungen der Alarme und der aktiven Sicherheitsfunktion auf die Leistung, Beanspruchung und das Situationsbewusstsein von Chirurgen untersucht. Die Ergebnisse zeigen, dass das NC-System sich positiv auf die Patientensicherheit sowie den physiologischen Aufwand von Chirurgen auswirkt und negativ auf die OP-Dauer und subjektive Beanspruchung, was vor allem auf die sehr invasive Abschaltfunktion zurück geht. Diese Erkenntnis hat zur Umgestaltung des Systems geführt: diese störende Abschaltfunktion wurde abgeschafft und stattdessen wird die Nähe der Risikostrukturen zusätzlich durch ein akustisches Alarm signalisiert. Eine weitere Untersuchung sollte klären, ob dadurch die negativen Auswirkungen reduziert werden können.

Chirurgische Navigationssysteme

Chirurgische Navigationssysteme sollen Chirurgen während einer Operation bei der räumlichen Orientierung im Situs unterstützen und somit zur Patientensicherheit beitragen. Mit einem spezifischen Zeigeelement, dem sogenannten Pointer, kann der Chirurg bei Bedarf anatomische Strukturen ansteuern. Auf der Grundlage der computer- oder magnetorsonanztomographischen Bildgebung und einer Trackingtechnologie wird der Pointer in Relation zu Patientenanatomie auf einem Navigationsbildschirm in drei Schnittebenen (coronal, sagittal und axial) dargestellt. Auf diese Weise kann der Chirurg die Lage des Pointers im Patient feststellen und die angesteuerte anatomische Struktur besser identifizieren. Diese Art der Naviga-

¹ Technische Universität Berlin, Zentrum Mensch-Maschine-Systeme, Graduiertenkolleg prometei

² Technische Universität Berlin, Institut für Psychologie und Arbeitswissenschaften

tion heißt pointer-basierte Navigation. Die Weiterentwicklung dieser Navigationssysteme bezieht sich auf das Tracking und die Darstellung der chirurgischen Instrumente wie Fräse, Shaver, Endoskop oder Sauger auf dem Navigationsbildschirm, so dass der Wechsel zum Pointer überflüssig wird.

Das neuere Navigated Control (NC)-System bietet zusätzlich zum Instrumenten-tracking einen optischen Alarm auf dem Navigationsbildschirm und eine aktive Schutzfunktion an. Dabei schaltet das NC-System die chirurgische Fräse beim Erreichen der prä-operativ definierten Grenzen des Arbeitsbereiches automatisch ab. Wenn das Instrument sich im Arbeitsbereich befindet, wird es auf dem Bildschirm grün dargestellt (Abb. 1). Sobald der Arbeitsbereich überschritten wird, wird das Instrument auf dem Bildschirm grau dargestellt und die Fräse wird abgeschaltet. Wenn das Instrument sich nicht im Blickfeld der Trackingkamera befindet, erscheint es rot und eine Meldung wird auf dem Navigationsbildschirm angezeigt. In diesem Fall hat das NC-System keine Information über die Lage des Fräse im Raum und somit keine Kontrolle über die Situation und schaltet die Fräse ebenfalls aus. Durch diese Maßnahmen unterstützt das NC-System aktiv den Chirurgen beim Schutz des Patienten gegen unbeabsichtigte Schädigungen.

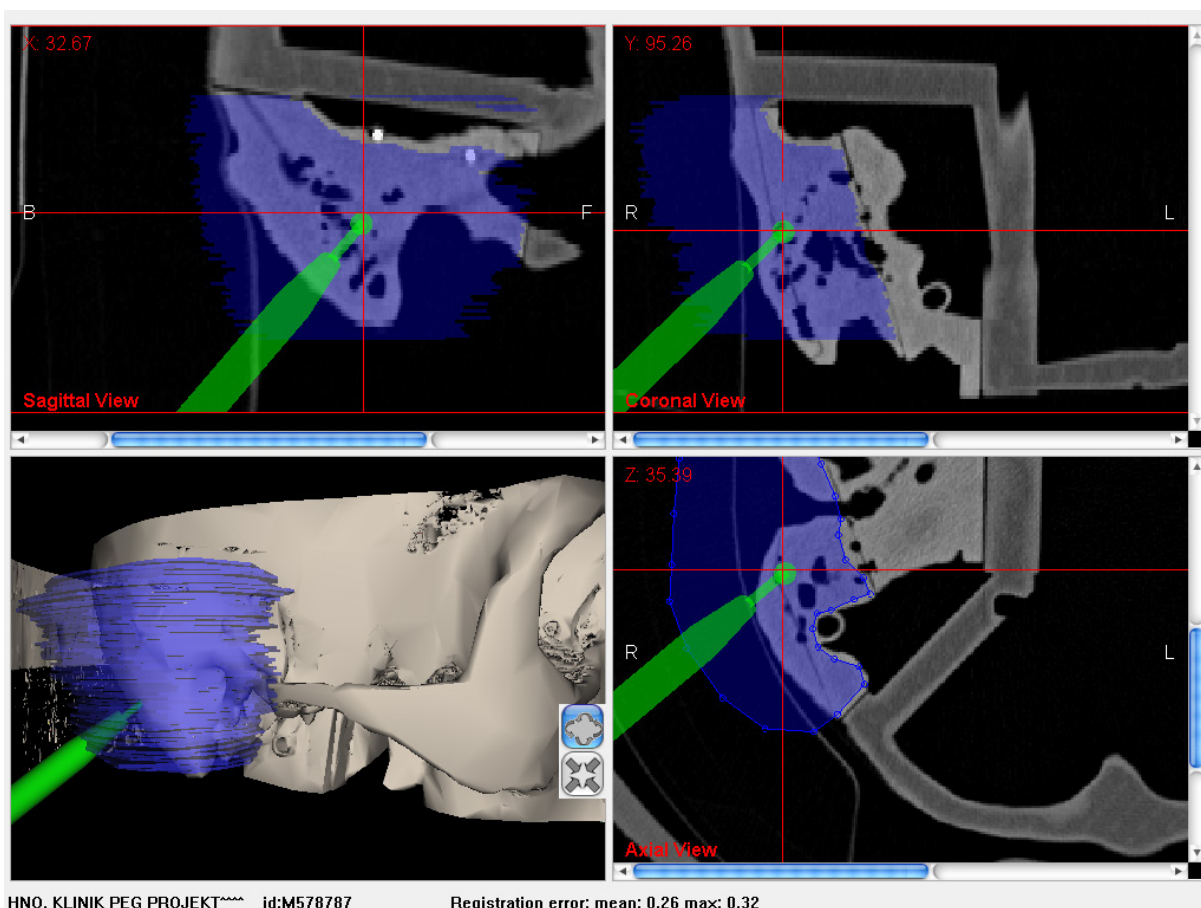


Abb.1: Screenshot des Navigationsbildschirms. Der Arbeitsbereich ist blau eingezeichnet. Das Instrument befindet sich innerhalb dieses Arbeitsbereiches und ist grün.

Rahmenmodell zur Klassifikation der Automationssysteme

Navigationssysteme können als Automation angesehen werden. Wobei die Orientierung im Situs und im Fall vom NC-System auch die Ausführung chirurgischer Eingriffe durch eine Maschine unterstützt wird (Manzey et al., 2009). Ein Rahmenmodell zur Klassifizierung unterschiedlicher Automationssysteme und zur Evaluierung von ihren Folgen hinsichtlich der menschlichen Leistung wurde von Parasuraman et al. (2000) vorgeschlagen. Demnach würden pointer-basierte Systeme ein relativ niedriges Automationsniveau darstellen, bei denen der Chirurg nur auf der Stufe der Informationsaufnahme und -analyse durch das System unterstützt wird. Das NC-System stellt ein automatisiertes System dar, bei dem durch die kontinuierlich getrackten chirurgischen Instrumente und die optischen Alarme die Stufen der Informationsaufnahme und -analyse höher automatisiert sind. Zudem sind die Stufen der Entscheidung und Handlungsausführung durch die Abschaltfunktion der Fräse teilweise automatisiert. Die Einordnung der beiden Navigationssysteme in das Rahmenmodell ist in der Abb. 2 dargestellt.

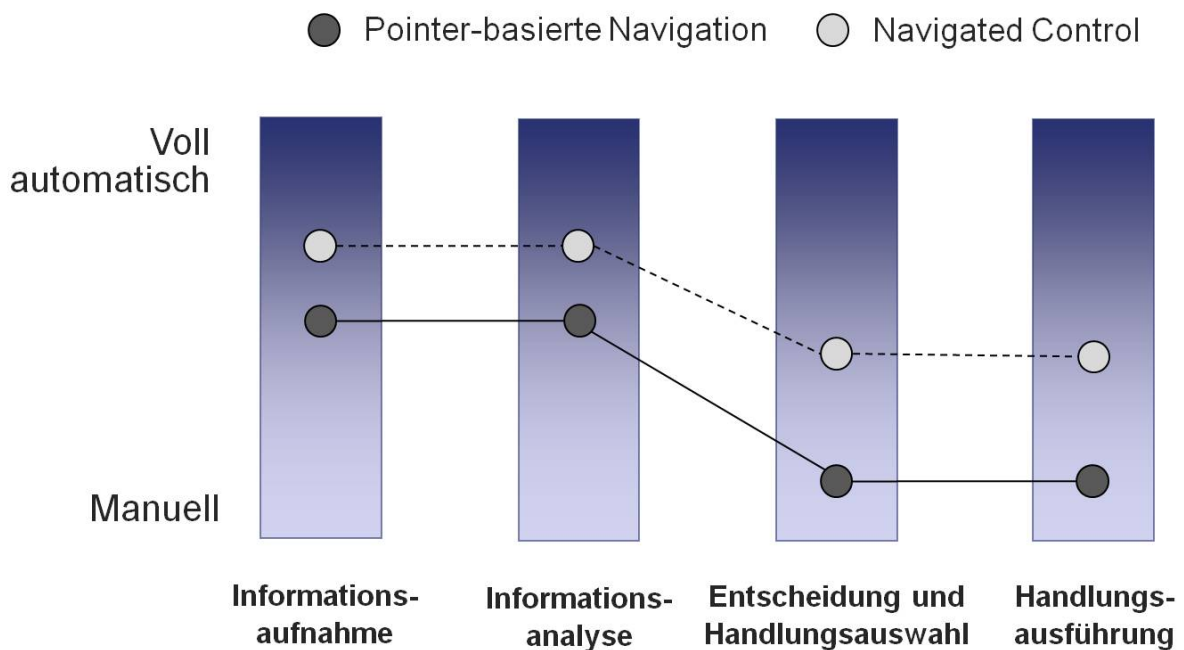


Abb.2: Klassifikation chirurgischer Navigationssysteme nach dem Model for Types and Levels of Human Interaction with Automation.

Das Ziel beim Einsatz von automatisierten Systemen ist neben der Leistungssteigerung die Fehlerreduktion, Erhöhung der Sicherheit sowie Reduktion von Belastung und Beanspruchung. Allerdings, wie zum Beispiel aus der Luftfahrt bekannt, können im Zusammenhang mit der Einführung der Automation neue Probleme und Risiken entstehen. Daher schlagen Parasuraman et al. (2000) folgende Evaluationskriterien der Automation vor: (i) Risiko des Übervertrauens in die Automation, (ii) erhöhte mentale Beanspruchung auf Grund der Notwendigkeit ein zusätzliches System zu bedienen, (iii) Schwierigkeit ein angemessenes Situations-

bewusstsein aufrecht zu erhalten auf Grund der mangelnden Einbindung in automatisierte Prozesse, (iv) Erwerb und Aufrechterhaltung der Fertigkeiten.

Methode

In zwei experimentellen Untersuchungen wurde der Einfluss des Einsatzes des NC-Systems auf die Leistung, sowie auf zwei der vier der von Parasuraman et al. (2000) vorgeschlagenen Kriterien-Beanspruchung und Situationsbewusstsein untersucht. Die Teilnehmer einer Untersuchung waren 14 fortgeschrittene Medizinstudierende oder Ärzte kurz nach Abschluss ihres Medizinstudiums (Novizenstudie). Für sie war die Nutzung des NC-Systems obligatorisch. An der zweiten Untersuchung nahmen sieben erfahrene HNO-Chirurgen teil, die das NC-System entsprechend dem klinischen Alltag frei einsetzten konnten (Expertenstudie). Das Modell für die Untersuchung war eine simulierte Mastoidektomie, eine anspruchsvolle Operation am Ohr. Die Teilnehmer führten die Mastoidektomie insgesamt 2 Mal durch: einmal mit und einmal ohne NC-System (d.h. manuell). Die Reihenfolge der Operationen wurde zwischen den Teilnehmern ausbalanciert. Da die Novizen noch keinerlei chirurgische Erfahrung besaßen, erhielten sie vor dem Versuch ein Training.

Als Leistungsparameter wurden gemessen 1) für die Durchführung der Mastoidektomie benötigte Zeit, 2) durch einen erfahrenen HNO-Chirurgen eingeschätzte Operationsqualität, 3) ebenfalls durch einen erfahrenen HNO-Chirurgen eingeschätzte mögliche post-operative Komplikationen und 4) Anzahl von unbeabsichtigten Verletzungen von sensiblen Strukturen. Die Erfassung der letzten Variablen erfolgte durch in dem simulierten Patienten integrierte Sensoren und dazugehörige Software.

Drei verschiedene Maße der Beanspruchung wurden in der Untersuchung erhoben: 1) subjektive Beanspruchung mit NASA-Task Load Index, 2) Blutdruck, Herzrate, Herzratenvariabilität und Respirationsrate als physiologischer Aufwand, 3) Leistung in der Sekundäraufgabe als Indikator für die Aufmerksamkeitsbindung. Die Sekundäraufgabe bestand darin, durch Betätigen eines Fußpedals so schnell wie möglich auf ein akustisches Signal zu reagieren. Gemessen wurden die Reaktionszeiten.

Nach einem bestimmten OP-Schritt wurden die Versuchsteilnehmer in ihrer Tätigkeit unterbrochen und zu ihrem Situationsbewusstsein mit Hilfe von SAGAT (Situation awareness global assessment technique, Endsley, 2000) befragt. Sie schätzten 1) die Entfernungen der verschiedenen Risikostrukturen zur Fräse, gaben 2) den derzeitigen Stand der Operation, 3) die Besonderheiten der Anatomie und 4) die verbleibende OP-Zeit an. Gemessen wurde die Richtigkeit ihrer Aussagen.

Ergebnisse

Die Ergebnisse beider Studien sind in der Tab. 1 dargestellt. Diese zeigen, dass NC-Systeme sich positiv auf die Patientensicherheit und physiologische Beanspruchung von Chirurgen auswirken. Auf der anderen Seite wirkt sich das NC-System negativ auf die OP-Dauer und die subjektive Beanspruchung der Chirurgen aus. Außerdem kommt es bei den Novizen zu Beeinträchtigungen der Leistung in der Sekundäraufgabe sowie zur weniger korrekten Beurteilung der anatomischen Gegebenheiten des Patienten. Ausführlich sind die Ergebnisse bei Manzey et al. (2011) und Luz et al. (2013) dargestellt.

Tab.1: Ergebnisse der Novizen- und Expertenstudie ("+" das NC-System hat im Vergleich zum manuellen Eingriff ein signifikant besseres Ergebnis geliefert; "-" das NC-System hat im Vergleich zum manuellen Eingriff ein signifikant schlechteres Ergebnis geliefert; "()" statistische Analyse war nicht möglich; "n.s." das Ergebnis war nicht signifikant.)

	Novizen	Experten
Leistung		
OP-Dauer	–	–
OP-Qualität	n.s.	+
Mögliche Komplikationen	+	(+)
Verletzungen	(+)	(–)
Beanspruchung		
Subjektiv	–	–
Physiologischer Aufwand		
Systolischer Blutdruck	+	+
Herzrate	+	n.s.
Herzratenvariabilität	n.s.	+
Respirationsrate	+	n.s.
Sekundäraufgabe	–	n.s.
Situationsbewusstsein		
Entfernungen	n.s.	n.s.
Stand der Operation	n.s.	n.s.
Anatomie	–	n.s.
Verbleibende Zeit	n.s.	n.s.

Schlussfolgerungen und Ausblick

Die negativen Auswirkungen werden vermutlich durch die sogenannte Line-of-Sight-Problematik verursacht, bei der das Instrument nicht nur im Fall der Nähe zu Risikostrukturen abgeschaltet wird, sondern auch, wenn die Sicht der Navigations-

kamera auf die Trackinginstrumente verdeckt wird, zum Beispiel durch das relativ große Mikroskop. Außerdem brachten die Untersuchungen die Erkenntnis, dass die Informationen auf dem Navigationsbildschirm kaum genutzt werden, da die Chirurgen ihren Blick vom Mikroskop abwenden müssen, um sie einzusehen.

Auf der Grundlage der Ergebnisse der beiden Untersuchungen, wurde ein neues System entwickelt (Distance Control) und das alte verbessert (Navigated Control 2). Bei beiden Systemvarianten wird die Entfernung zum segmentierten Bereich angezeigt und beim Erreichen der Grenzen des segmentierten Bereiches werden ein optischer und zusätzlich ein akustischer Alarm ausgelöst. Bei Navigated Control 2 bleibt die Abschaltfunktion bestehen, es wird jedoch dem Chirurgen die Möglichkeit angeboten, diese mit deutlich reduzierter Drehzahl zu überschreiben.

Um die negativen Folgen der Line-of-Sight-Problematik zu reduzieren, wird das große Mikroskop durch einen schlanken Exoskop ersetzt. Dadurch werden auch die Informationen auf dem Navigationsbildschirm, der neben dem Exoskopbildschirm stehen, besser für den Chirurgen verfügbar.

Eine weitere Untersuchung dient dazu, zu überprüfen, ob die negativen Auswirkungen des NC-Systems durch die vorgenommenen Veränderungen reduziert werden können, die Vorteile jedoch erhalten bleiben. In dieser Untersuchung führten 18 HNO-Assistenz- und -Fachärzte drei Mal eine Mastoidektomie durch: einmal ohne Assistenzsystem (Kontrollbedingung), einmal mit Distance Control und einmal mit Navigated Control 2. Die Ergebnisse werden bei der Berliner Werkstatt Mensch-Maschine-Systeme vorgestellt.

Literatur

- Endsley, M. (2000). Direct measurement of SA: Validity and use of SAGAT. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness: Analysis and measurement* (pp. 147-173). Mahwah, NJ: Lawrence Erlbaum.
- Luz, M., Manzey, D., Mueller, S., Dietz, A., Meixensberger, J., & Strauss, G. (2013). Impact of navigated-control assistance on performance, workload and situation awareness of experienced surgeons performing a simulated mastoidectomy. *The international Journal of Medical Robotics and Computer Assisted Surgery*. doi: 10.1002/rcs.1527
- Manzey, D., Luz, M., Mueller, S., Dietz, A., Meixensberger, J., & Strauss, G. (2011). Automation in surgery: the impact of navigated-control assistance on performance, workload, situation awareness, and acquisition of surgical skills. *Human Factors*, 53, 584-599.
- Manzey, D., Strauss, G., Trantakis, C., Lueth, T., Roettger, S., Bahner-Heyne, J. E., Dietz, A., & Meixensberger, J. (2009). Automation in surgery: A systematic approach. *Surgical Technology International*, 18, 37-45.

Parasuraman, R., Sheridan, T., & Wickens, C. D. (2000). A model of types and levels of human interaction with automation. *IEEE Transactions of Systems, Man, and Cybernetics. Part A—Systems and Humans*, 30, 286-297.

Der Einfluss von Immersivität in virtuellen Umgebungen auf Situation Awareness im Fahrzeug

Diana Reich¹

Schlüsselwörter: Immersivität, virtuelle Umgebung, Situation Awareness, Cockpit-Interaktionen

Zusammenfassung

Innovative Fahrzeug-Cockpitentwicklungen führen zu Unterschieden hinsichtlich Fahrperformanz, sowie Risiko- und Situationswahrnehmung während typischer Dual-Task-Aufgaben im Fahrzeug (z. B. Navigationseingaben). Mittels prospektiver Evaluation der Fahrerfähigkeiten und Bedürfnisse, sowie den darauf ausgelegten Cockpit-Prototypen, ließen sich Verkehrssicherheit steigern und Entwicklungskosten reduzieren. Zusätzlich können durch virtuell projizierte, immersive Versuchsumgebungen realistischere Dual-Task-Szenarien in sehr frühen Entwicklungsstadien evaluiert werden. In verschiedenen Untersuchungsschritten soll systematisch untersucht werden, welchen Einfluss Immersivität dabei auf die prospektive Evaluation von Mensch-Cockpit-Interaktionen hat, um reliable und valide Aussagen treffen zu können. Die ersten Versuche haben gezeigt, dass die Shutterbrille keinen Einfluss auf die Performanz oder subjektive Wahrnehmung der Probanden hat. Dadurch sollte deren Verwendung innerhalb immersiver Versuchsumgebungen keine Verzerrung der Ergebnisse mit sich bringen. Weiterhin konnte ein positiver Einfluss von Immersivität auf die Performanz hinsichtlich der Bearbeitungszeiten mit dem Regler beobachtet werden. Abschließend werden Folgeexperimente vorgestellt, welche die eigentliche Primäraufgabe "Fahren" integrieren und den Einfluss immersiver Versuchsumgebungen auf Situation Awareness untersuchen.

Einleitung

Durch eine prospektive Evaluation von Fähigkeiten und Bedürfnissen zukünftiger Nutzer (Fahrer), sowie den darauf ausgelegten Cockpit-Prototypen, ließen sich nicht nur Verkehrssicherheit und Fahrerzufriedenheit steigern, sondern auch spätere Servicekosten und mögliche Imageschäden reduzieren (Mayhew, 1999). Allerdings sind die dafür notwendigen iterativen Verfahren bisher noch mit einem hohen zeitlichen und finanziellen Aufwand verbunden, vor allem im Falle von notwendigen Neukonstruktionen.

In der Automobilindustrie sind beispielsweise verschiedene Strömungen im Bereich der Cockpitentwicklung zu beobachten (Ablaßmeier, 2008). Diese unterscheiden sich vor allem in der Art und Weise der Interaktion mit den Bedienele-

¹ Graduiertenkolleg prometei, Marchstr. 21, 10587 Berlin, dreich@zmms.tu-berlin.de

menten (Bubb, 2008a). Die verschiedenartigen Dual-Task-Aufgaben (Fahren und Bedienen) im Fahrzeug führen in Folge der verschiedenen Interaktionskonzepte zu Unterschieden hinsichtlich Performanz, Risikowahrnehmung und Situation Awareness (z. B. Baumann, Petzoldt & Krems, 2006). Das übergeordnete Ziel bei der Entwicklung innovativer Fahrzeugcockpits ist die Gewährleistung von Sicherheit während der Primäraufgabe Fahren. Um dieses Ziel zu erreichen sollten Standards entwickelt und Methoden evaluiert werden, welche eine reliable und valide Abbildung der Mensch-Cockpit-Interaktion in frühen Phasen der Entwicklung ermöglichen.

Durch den Einsatz simpler Prototypen innerhalb virtueller, immersiver bzw. hybrider (=Kombination aus virtuellen und physischen Komponenten) Testumgebungen würden nicht nur die anfallenden Kosten in der Fahrzeug-Cockpitentwicklung minimiert, sondern vor allem auch eine realistischere Evaluation der Mensch-Cockpit-Interaktion in frühen Entwicklungsstadien realisiert werden. Mittels Anpassung bereits bestehender Daten in der jeweiligen (Konstruktions-) Software ist es möglich sowohl Fahrzeug, Innenraum und Cockpit, als auch die Interaktions- bzw. Fahrumgebung jederzeit und ohne großen Mehraufwand zu verändern und erneut zu evaluieren. Beispielsweise könnte ein Touch-Bediensystem im Kleinwagen auf der Landstraße, ein Transporter im Stadtverkehr oder eine Dreh-Drücksteller-Variante erprobt werden. Demnach könnten durch den Einsatz virtueller Fahrzeugs- und Umgebungsprojektionen deutlich realere bzw. kritischere Verkehrsszenarien in frühen Phasen der Produktentwicklung zuverlässig abgebildet werden. Dadurch ließen sich zentrale Konstrukte wie Risikowahrnehmung und Situation Awareness während prototypischer Dual-Task-Aufgaben im Fahrzeugcockpit reliabel und valide evaluieren.

Methodik

Im ersten Untersuchungsschritt wurden zentrale Einflussgrößen innerhalb immersiver (virtueller) Umgebungen untersucht. Anhand eines selbst entwickelten Navigationssystem-Interfaces (NAVE = Navigation system Virtual Environment) wurde überprüft, ob und welchen Einfluss (A) die Shutterbrille¹ und (B) ein virtuell projiziertes Fahrzeug innerhalb drei unterschiedlicher Untersuchungsumgebungen (nicht-immersiv ohne Shutterbrille (1), nicht-immersiv mit Shutterbrille (2) und immersiv mit Shutterbrille (3)) haben. Die Aufgabe der Probanden bestand jeweils in der Eingabe von Zielen in das Navigationssystem NAVE ohne Fahren. Dieser Untersuchungsschritt bestand aus zwei Experimenten und wurde einerseits mit Touch-Eingaben (N = 20) und zum anderen mit (Dreh-) Regler-Eingaben (N = 20) durchgeführt. Anschließend wurden die unterschiedlichen Interaktionseingaben (Touch, Regler) ebenfalls miteinander verglichen. Es wurden sowohl objektive Performanzmaße (Zeit, Fehler), als auch subjektive Daten (AttrakDiff) der Versuchspersonen erhoben.

¹ Die Shutterbrille wird zum Sehen von stereoskopischen (3D) Bildern benötigt.

Ergebnisse

Um die Unterschiede der Einflussgrößen (A und B) zu überprüfen wurde eine ANOVA mit Messwiederholung durchgeführt, wobei die drei Testumgebungen (Bedingung 1, 2, 3 siehe Tabelle 1) als unabhängige Variable fungierten.

Tab. 1: Bearbeitungszeit in s und Anzahl der Fehler in allen drei Bedingungen

		Bedingung 1	Bedingung 2	Bedingung 3
		nicht-immersiv; ohne Shutterbrille	nicht-immersiv; mit Shutterbrille	immersiv; mit Shutterbrille
Touch	Zeit	223s	235s	224s
	Fehler	2,9	2,6	2,5
Regler	Zeit	505s*	499s	451s*
	Fehler	2,7	2,8	2,3

*Signifikante Unterschiede

Die Ergebnisse zeigen, dass die Shutterbrille keinen Einfluss auf die Aufgabebearbeitung hat (Bedingung 1 vs. Bedingung 2). Allerdings zeigte sich ein signifikanter Unterschied hinsichtlich des virtuell projizierten Fahrzeugs (immersiv; Bedingung 3), im Vergleich zu der Testbedingung ohne Projektion (nicht-immersiv; Bedingung 1) für die Interaktion mit dem Regler: ($F(1,19) = 4.81$, $p < 0.005$, $\eta^2_{\text{part}} = 0.202$). Hier waren die Bearbeitungszeiten signifikant schneller innerhalb der immersiven Testbedingung mit Projektion, verglichen mit der nicht-immersiven Bedingung ohne Projektion (siehe Abbildung 1). Dieser Effekt wurde innerhalb der Touch-Bedingung jedoch nicht bestätigt.

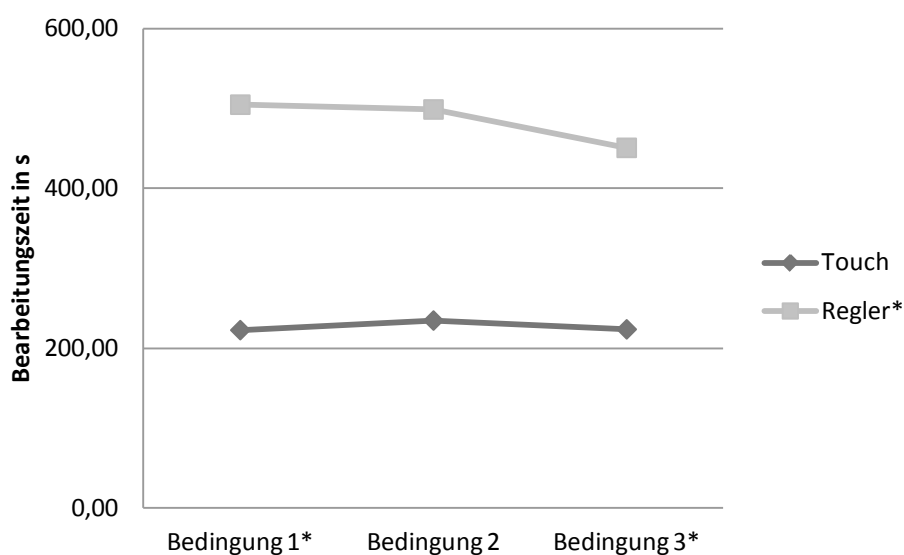


Abb.1: Navigationseingaben (Touch, Regler) innerhalb unterschiedlicher Bedingungen

Es wurde kein signifikanter Unterschied bei der Fehleranzahl hinsichtlich der Versuchsumgebung (Bedingung 1, 2, 3) und/oder der Eingabeart (Touch, Regler) gefunden. Allerdings zeigen die absoluten Fehlerzahlen bei beiden Eingabearten, dass sich die immersive Umgebung (Bedingung 3) auch hier tendenziell positiv auf die Performanz auswirkt, im Vergleich zu den beiden nicht-immersiven Bedingungen 1 und 2 (siehe Tabelle 1). Weitere Untersuchungen sind notwendig, um zuverlässige Aussagen treffen zu können. Die Analyse der subjektiven Daten (Attrak-Diff) ergab keine signifikanten Unterschiede. Die Versuchspersonen haben in diesen Experimenten keinen Unterschied zwischen immersiver bzw. nicht-immersiver Umgebung bei der Interaktion mit verschiedenen Eingabegeräten wahrgenommen.

Diskussion

Die Ergebnisse des ersten Untersuchungsschrittes haben gezeigt, dass die Shutterbrille keinen Einfluss auf die Performanz oder subjektive Wahrnehmung hat. Die virtuelle Projektion hingegen führte im Vergleich zu der Testbedingung ohne Projektion zu signifikant schnelleren Bearbeitungszeiten. Dieses Ergebnis lässt vermuten, dass sich immersive Testumgebungen positiv auf das Performanzverhalten der Nutzer auswirken, wenn sie mit dem Regler interagieren. Dass dieser Effekt innerhalb des Touch-Experiments nicht bestätigt werden konnte, liegt möglicherweise an dem sehr schnellen Anfangsniveau der Touch-Eingaben, welche eine Steigerung unter Immersivität nur schwer realisierbar machten. Der positive Einfluss von Immersivität auf die Performanz hinsichtlich Fehleranzahl konnte nicht abschließend bestätigt werden. Wie bereits erwähnt sind weitere Untersuchungen notwendig, um mögliche signifikante Effekte hinsichtlich der Fehlerrate unter unterschiedlichen Testbedingungen zu finden bzw. einen Zufall auszuschließen. Überraschenderweise ergab die Analyse der subjektiven Daten ebenfalls keine signifikanten Unterschiede. Versuchspersonen konnten in den Versuchen keinen Unterschied zwischen immersiver bzw. nicht-immersiver Umgebung bei der Interaktion mit verschiedenen Eingabegeräten wahrnehmen. Dieses Ergebnis könnte damit erklärt werden, dass die virtuelle Projektion in keinem direkten Zusammenhang mit der Aufgabenbearbeitung stand. Würde die immersive Umgebung, hier die Fahrzeugs- und/oder Umgebungsprojektion, eine Performanz bezogene Rolle spielen (z. B. beim Fahren und interagieren), verändert sich wahrscheinlich auch das subjektive Empfinden der Probanden. Dementsprechend ist die Integration der Primäraufgabe "Fahren" essentiell notwendig, um zuverlässige Aussagen zu treffen.

Ausblick

In einer aktuell laufenden Untersuchung wird die eigentliche Primäraufgabe "Fahren" integriert (siehe Abbildung 2, links). In diesem Fahrversuch soll überprüft werden, ob ein zunehmender Grad an Immersion (3D vs. 2D; ohne Ton vs. mit Ton) das Fahrverhalten (Performanz) und Fahrerleben (Situation Awareness) be-

einflusst. Um eine stereoskopische bzw. dreidimensionale Fahrumgebung zu realisieren wurde eine Holobench verwendet (siehe Abbildung 2, rechts). In diesem Zusammenhang werden folgende Messdaten erhoben: (A) objektive Fahrmaße des standardisierten Lane Change Tasks (LCT), (B) Blickdaten, (C) subjektive Eindrücke zum Immersions- und Präsenzerleben (nach Scheuchnpflug, 2001) und (D) die subjektiv wahrgenommene Beanspruchung (Nasa TLX). Zusätzlich wird das SAGAT-Verfahren (Situation Awareness Global Assessment Technique; nach Endsley, 1995a) zur Überprüfung der aktuellen Situation Awareness während unterschiedlicher Fahrbedingungen eingesetzt. Zum aktuellen Zeitpunkt liegen noch keine Ergebnisse vor.



Abb.2: Probandin im Fahrsimulator (links) und LCT auf der Holobench (rechts)

In Planung befindet sich ein weiterer Versuch, welcher die wenig immersive Testumgebung (2D; ohne Ton) und die hoch immersive Testumgebung (3D; mit Ton) mit den drei unterschiedlichen prototypischen Interaktions-Eingaben (Touch, Regler, Geste) vergleichen soll. Zum einen interessiert hierbei erneut der Unterschied bzw. Einfluss verschiedenartiger Immersivitätsstufen auf das Fahrverhalten- und Erleben, zum anderen aber auch, ob und welchen Einfluss unterschiedliche Eingabemodalitäten in diesem Zusammenhang haben.

Literatur

- Ablaßmeier, M. (2008). Multimodales, kontextadaptives Informationsmanagement *im Automobil*. Dissertation. Technische Universität München.
- Bubb, H. (2008). Produktergonomie _ Skriptum zur Vorlesung. München : Lehrstuhl für Ergonomie.
- Baumann, M. R. K., Petzoldt, T., & Krems, J. F. (2006). Situation Awareness beim Autofahren als Verstehensprozess. *MMI-Interaktiv*, 11, 43-57.
- Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. *Human Factors* 37, 65-84.

- Mayhew, D. J. (1999). *The Usability Engineering Lifecycle*. San Francisco: Morgan-Kaufmann.
- Scheuchpflug, R. (2001). Measuring presence in virtual environments. In M.J. Smith, G. Salvendy, & M. R. Kasdorf (Eds.). *HCI International 2001*, pp. 56-58. New Orleans: HCI International.

Klassifikation und Detektion des Nutzungskontextes bei Smartphones

Ralf Reichmuth¹, Robert Schleicher², Sebastian Möller²

Schlüsselwörter: Klassifikation des Nutzungskontextes, mobiler Nutzungskontext, Smartphone, Detektion des Nutzungskontextes

Zusammenfassung

Mobile Anwendungen auf Smartphones werden in den unterschiedlichsten Nutzungskontexten verwendet. Um die Komplexität der Anwendungen für den Nutzer zu reduzieren wächst die Bedeutung von kontextsensitivem Verhalten dieser Anwendungen. Hierzu ist eine genaue Beschreibung und Klassifikation des Nutzungskontextes notwendig, die bislang noch nicht verfügbar ist.

Diese Arbeit soll helfen diese Lücke zu schließen. Dies geschieht einerseits durch die Vorstellung eines Modells des Nutzungskontextes bei Smartphones, zum anderen durch eine empirische Studie. Das Modell teilt den Nutzungskontext in einzelne Dimensionen auf. Im Rahmen einer Feldstudie wird der Nutzer angehalten, seinen Nutzungskontext direkt mittels eines Smartphone-Fragebogens zu beschreiben, wobei gleichzeitig Sensor- und Nutzungsdaten aufgezeichnet werden. Durch diese Feldstudie kann der Nutzungskontext und dessen Bedeutung erfasst werden, sowie Sensor- und Nutzungsdaten diesem Kontext zugeordnet werden.

Einleitung

Mobile Geräte wie Smartphones haben sich innerhalb kurzer Zeit in unser tägliches Leben integriert. Smartphones beeinflussen die Kommunikation und das gesellschaftliche Leben selbst dann, wenn wir persönlich kein solches Gerät besitzen. Sie werden in den unterschiedlichsten Kontexten verwendet, wobei es bisher noch kein einheitliches Modell und nur unzureichende Forschung zur Bestimmung des Nutzungskontexts gibt. Diese Arbeit soll helfen, diese Lücke zu schließen.

Modell des Nutzungskontextes

Für das zu bestimmende Modell wird die Definition von Kontext von Dey (2001) benutzt. Dieses besagt: „Kontext ist jede Information, die dazu genutzt werden kann, um die Situation einer Entität zu beschreiben. Eine Entität ist hierbei eine Person, Ort oder Objekt, welches als relevant betrachtet wird für die Interaktion zwischen Nutzer und Anwendung, eingeschlossen den Nutzer und die Anwendung selber.“ In der Literatur gibt es viele Modelle, die den Nutzungskontext mobiler

¹ Zentrum Mensch-Maschine-Systeme, GRK prometei, TU Berlin, Germany

² Quality and Usability Lab, Telekom Innovation Laboratories, TU Berlin, Germany

Geräte in einzelne Dimensionen unterteilen. Es werden daher einige dieser Unterteilungen kurz vorgestellt. Schilit et al. (1994) beschrieb den Nutzungskontext bei kontextsensitiven Geräten als „Wo du bist, mit wem du zusammen bist und welche Ressourcen in der Nähe sind.“. Lee et al. (2005) unterteilt den Nutzungskontext für das mobile Internet in „Erregung“, „Vergnügen“, „Arbeitszeit“, „Zeitdruck“, „Bewegung“, „Ort“, „Ablenkung“, „Gedränge“, „Interaktion“ und „Datenschutz“. Des Weiteren existieren viele spezielle Beschreibungen des Nutzungskontextes, welche aber kaum eine Verallgemeinerung zulassen.

Zur Bestimmung des Nutzungskontextes bei Smartphones wird in diesem Beitrag das Arbeitsmodell von Schmidt (1999) verwendet. Dieses Modell bezieht sich auf mobile Computer und wird in dieser Arbeit um die Dimension „Zeit“ erweitert (Abbildung 1).

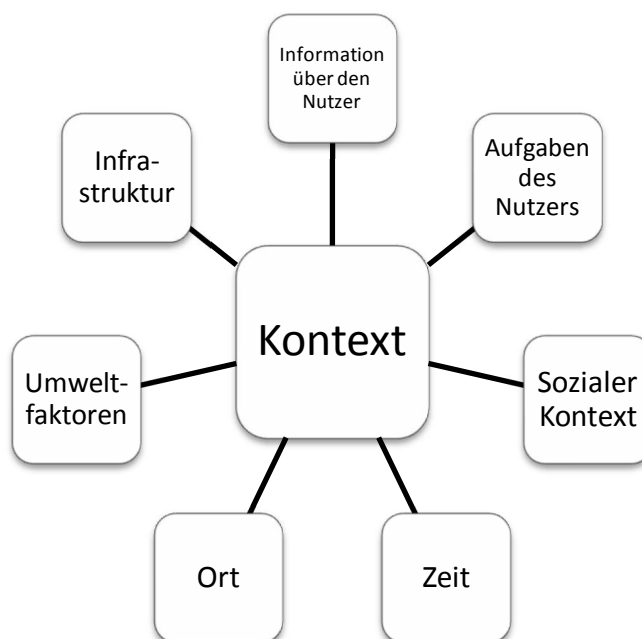


Abb.1: Modell des Nutzungskontextes.

Diese Kontextdimensionen wurden aufgeteilt auf die menschlichen Faktoren (Informationen über den Nutzer, Aufgaben des Nutzers und sozialer Kontext), die physikalische Umgebung (Infrastruktur, Umweltfaktoren und Ort) und die Zeit. Diese Dimensionen können im Anschluss noch weiter differenziert werden.

Mit Hilfe der geplanten Studie soll unter anderem überprüft werden, ob dieses Modell vollständig ist. Der Proband wird hierbei am Ende des Fragebogens gefragt, ob ihn abgesehen von diesen Dimensionen noch etwas Weiteres beeinflusst hat.

Studie

Durch eine explorative Studie soll versucht werden Nutzungskontexte zu bestimmen, diesen Nutzungskontexten Bedeutungen zuzuweisen, sowie Sensor- und Nutzungsdaten zu Nutzungskontexten zuzuordnen.

Methode

Im Rahmen dieser Arbeit wurde eine native iPhone-App entwickelt. Diese App wurde auf das private Handy von 17 Probanden aufgespielt. Jeder Proband wurde vom Versuchsleiter instruiert, an zwei Tagen die Kontexte jedweder Nutzung zu dokumentieren. Hierzu konnten die Probanden jeweils einen Wochentag (zwischen Montag und Freitag) und einen Freizeit-Tag (zwischen Samstag und Sonntag) für die Dokumentation auswählen. Die Probanden sollten jedes Mal vor einer Smartphone-Nutzung den Nutzungskontext in der App protokollieren. Hierzu wurden die Benachrichtigungen der App genutzt, um direkt vom gesperrten Bildschirm in einen Fragebogen zu springen, welcher den Nutzungskontext beschreibt. Die Probanden wurden zusätzlich unterrichtet, dass im Hintergrund kontinuierlich Sensor- und Nutzungsdaten aufgezeichnet werden.

Die Bedienung der App wurde hierbei so gestaltet, dass der Proband möglichst schnell und flexibel seinen Kontext eingeben kann, die Studie nicht zu belastend ist und die normale Interaktion möglichst wenig beeinflusst wird. Jedem Probanden wurden 15 Fragen (14 verpflichtende und eine optionale Frage) zu seinem Nutzungskontext gestellt. Auf jedem Bildschirm der App wird eine Frage dargestellt. Sobald eine Antwortmöglichkeit ausgewählt wird gelangt der Proband automatisch zur nächsten Frage.

Der gesamte Fragebogen wurde hierbei nach dem Arbeitsmodell von Schmidt (1999) ausgerichtet, so wurden Fragen erstellt, die die einzelnen Dimensionen des Nutzungskontexts abdecken. Zusätzlich wurde noch die Bedeutung des Nutzungskontexts abgefragt und ob der Proband von weiteren Kontextinformationen beeinflusst wurde. Tabelle 1 veranschaulicht den Aufbau des Fragebogens, dessen Beantwortung in einer vorangestellten Pilotstudie im Durchschnitt 52 Sekunden dauerte.

Tab.1: Aufbau des Fragebogens.

Kategorie der Frage	Anzahl der Fragen
Zeitpunkt der Interaktion	1
Bedeutung des Nutzungskontexts	2
Emotionen des Probanden	3
Aufgaben des Probanden	3 + 1(optional)
Sozialer Kontext	1
Ort	1
Umwelteinflüsse	1
Infrastruktur	1
Weitere Kontextinformationen	1

Zusätzlich zu der manuellen Eingabe des Nutzungskontextes vom Probanden läuft die App während der Studie im Hintergrund und zeichnet Daten über das Nutzerverhalten, sowie den Zustand der App und des zugehörigen Gerätes auf. Diese Daten lassen sich in drei Kategorien einteilen: gerätespezifischen Informationen, Sensor- und Nutzungsdaten und Ereignisse. Die gerätespezifischen Informationen werden nur einmalig bestimmt. Die Sensor- und Nutzungsdaten sind Daten wie der Ort (über GPS), die Lautstärke etwaiger Umgebungsgeräusche (über das Mikrophon), die Bewegung des Gerätes (über Beschleunigungssensor), der Zeitpunkt der Interaktion, Verbindungsinformationen, laufende Prozesse sowie Informationen zum Ladezustand des Akkus. Diese Daten werden beim Start der Anwendung, beim Schließen der Anwendung sowie alle zehn Minuten bestimmt. Im Unterschied dazu treten die Ereignisse zu beliebigen Zeitpunkten auf. Ereignisse sind zum Beispiel das Erhalten einer SMS, eines Anrufs und Veränderung in der Feldstärke des GSM-Empfängers.

Die ausgewählten Antwortmöglichkeiten und die im Hintergrund gespeicherten Daten werden anschließend an einen Server der Technischen Universität Berlin zur weiteren Auswertung gesendet.

Ergebnisse

Die App wurde an 17 Probanden verteilt. Hieraus sollen wir relevante und häufig auftretende Nutzungskontexte klassifiziert werden. Im Rahmen dieser Arbeit wurde eine Pilotstudie mit vier Probanden durchgeführt. Erste Resultate sind hierbei vielversprechend, da bereits Cluster von Nutzungskontexten gebildet werden konnten, welche für einen Großteil der Interaktionen verantwortlich waren. Diese Resultate stimmen mit der Studie von Lee et al. (2005) überein, bei der drei Nutzungskontexte für 24% der verursachten Session verantwortlich waren. Des Weiteren legt die Pilotstudie nahe, dass Smartphones oft im stationären Kontext benutzt werden und eher selten im mobilen Kontext. So wurden in der Pilotstudie nur 11% der Nutzungen im mobilen Kontext aufgezeichnet und 57% zu Hause. Diese Resultate stimmen mit der Studie von Verkasalo (2008) überein, bei der 27,1% der aktiven Smartphone-Nutzungszeit im mobilen Kontext und 51,8% zu Hause stattfanden.

Ausblick

Die Ergebnisse dieser Studie sollen anschließend genutzt werden, um Nutzungskontexte und deren Relevanz mit Sensor- und Nutzungsdaten zu korrelieren. In der nahen Zukunft können die Ergebnisse dieser Arbeit dazu genutzt werden, die Interaktion mit Smartphones zu verbessern, indem z. B. nicht relevante Informationen für Nutzer ausgeblendet werden und bestimmte Funktionen je nach Nutzungskontext automatisiert ausgeführt werden.

Literatur

- Lee, I., Kim, J., & Kim, J. (2005). Use Contexts for the Mobile Internet: A Longitudinal Study Monitoring Actual Use of Mobile Internet Services. *International Journal of Human-Computer Interaction*, 18(4), 269–292.
- Schilit, B., Adams, N., & Want, R. (1994). Context-aware computing applications. In: *Proceedings of the 1994 First Workshop on Mobile Computing Systems and Applications*, IEEE Computer Society, Washington, DC, USA, 85–90.
- Schmidt, A., Beigl, M., & Gellersen, H.-W. (1999). There is more to context than location. *Computers & Graphics*, 23(6), 893–901.
- Verkasalo, H. (2008). Contextual patterns in mobile service usage. *Personal and Ubiquitous Computing*, 13(5), 331–342.

Collaboration Enable Techniques for Complex Enterprise Resource Planning Projects

Taru Salmimaa¹

Keywords: Enterprise Resource Planning (ERP), Knowledge Sharing, Participatory Design (PD), Root Cause Analysis, System Dynamics

Abstract

Stakeholders of Enterprise Resource Planning (ERP) system implementation aim to understand each other and want to collaborate effectively. More effort is needed to make sense what is a right timing and focus of stakeholders' involvement. An ERP system as being modular and commercial product is typically critical for business. With an ERP a company aims to bring a value by streamlining business processes. Earlier studies prove that the Participatory Design (PD) approach enhances knowledge sharing between stakeholders in the implementation of ERP systems. This paper describes some additional techniques from complex system analyses which can be included to the traditional PD practices. Outcome of these techniques will be explored in a future case study research. Complex system analyzing techniques such as System Dynamics can support a more holistic view of the project environment. An ERP project is considered as a complex system because interactions of participant stakeholders and processes in a specific project environment cannot be fully pre-planned.

Introduction

Enterprise Resource Planning (ERP) system as being a modular software product for business information and operations management integrates the business functions of an enterprise (Davenport, 1998). Current ERP systems aim to cover the functionality across the organizational boundaries with supply-chain management and extranets of the external users. It can be challenging to generalize a scope of ERP system. For example, it can cover the whole value chain of an organization including participants from partners and consumers (Koh, Gunasekaran, & Goodman, 2011), (Damsgaard & Karlsbjerg, 2010).

An ERP project typically needs to combine the right knowledge of ERP development, systems integration, requirements management, business processes re-engineering and project management in inter-organizational relationships (Wright & Wright, 2002) for proceeding in the implementation tasks as planned. Commitment and collaboration of all stakeholders are necessary for achieving the trust and fluent communication throughout the implementation project. Participants have

¹ Tampere University of Technology
Department of Information Management and Logistics

typically predefined roles which have different amount of decision making power and communication interfaces depending on the roles' centrality for the implementation. For example, top manager (steering group), mid-manager (ERP project manager), informant (users, business owners) or specialist (consultant, ERP developers and architects) are all very important roles to be involved in the ERP system.

Participatory Design (PD) approaches have been applied to enhance the collaboration and knowledge integration within the ERP implementation when the participants come from different backgrounds and areas of knowledge (Pries-Heje & Dittrich, 2009). It has been studied that PD techniques lower the hierarchical structures between the participants who have different role in the project (Wagner & Piccoli, 2007). Having a shared view of the desired system functionality is sometimes challenging why more communities of practice and centers of knowledge integration in the large ERP implementation project are also needed (Pawlowski & Raven, 2000; Pries-Heje & Dittrich, 2009; Wu & Cao, 2009).

Socio-Technical System Thinking

In this paper an ERP implementation project is considered as a socio-technical system (Trist, 1981). Socio-technical system approach which has been developed by Tavistock Institute in London has been developed to improve job satisfaction in British coal mining industry (Trist, 1981). Recently socio-technical approach has been used in Information System (IS) research for explaining the change process within an IS implementation (Ahmad, Lyytinen, & Newman, 2011; Mumford, 2000). Having an understanding how the theory can be put into practice the socio-technical system approach is an example in which researchers have developed plural practical methods such as participatory design (Kensing & Blomberg, 1998) and soft system methodology (Checkland, 1981) for making the approach more a practical process. For example, many User-Centered Design (UCD) methods are based on the socio-technical system thinking and have become popular in the large extent of software design practices (Beyer & Holtzblatt, 1998).

According to Forrester (1971, p. 111) "We do not know enough to design the most effective social systems directly without first going through a model-building experimental phase". He was a pioneer who talked about mental models of participants which should be analyzed for making more sense of the social system. He also empirically tested and proved that a behavior of social system is hard to predict (Forrester, 1971). If the system is complex as the most of ERP implementation projects are the visualization of the system overall behavior throughout the system lifecycle certainly supports a better decision-making in unpredictable ERP project environments.

Collaboration Enable Techniques for ERP Projects

The following paragraphs describe steps which can be part of a PD process within an inter-organizational ERP implementation. The purpose is to share an understanding among stakeholders of the complex and dynamic nature of an ERP project through the system that is implemented.

Step 1: Visualization of the Socio-Technical System

Figure 3 represents four accumulations and activities which are identified in the development process of a socio-technical system. This diagram demonstrates how interactions are formed. According to the case studies of Luna-Reyes, Zhang, Ramón Gil-García, & Cresswell (2005) all the four accumulations are changed during the projects: *system requirements*, *organizational design*, *system functionality* and *knowledge about practice*. Within this study it is justified how the feedback loops are consistent with a theory. For example, by applying the structuration theory of (Giddens, 1984) the authors argument “*The inner process, linking knowledge about practice with organizational design is consistent with the concept of the duality of the structure (Giddens, 1984)... and the process in the outside, linking knowledge about practice with system requirements and system functionalities is consistent with the concept of the duality of technology (DeSanctis & Poole, 1994; W. J. Orlikowski, 1992; Wanda J Orlikowski, 1996)*”. As a result, the visualization of the accumulations and the variables which change the state of accumulations can support an understanding what is a dynamic nature of requirements management in the complex system development.

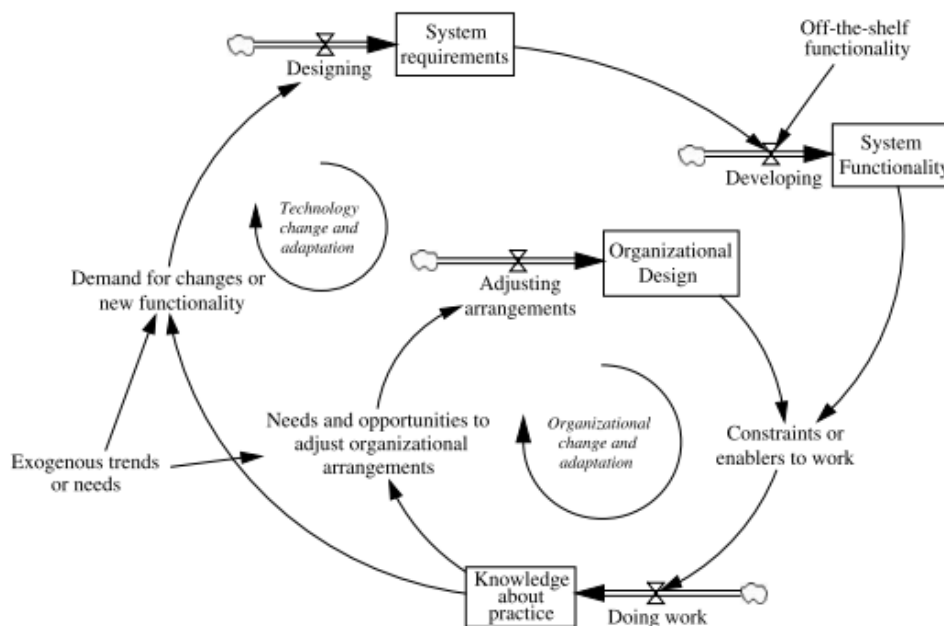


Figure 3: Recursive interactions representing the theoretical constructs of the duality of structure and the duality of technology (Luna-Reyes, Zhang, Ramón Gil-García, & Cresswell, 2005).

System dynamics approach has a different ground from the socio-technical system thinking. But as pointed out in the case study of Luna-Reyes, Zhang, Ramón Gil-García, & Cresswell (2005) system dynamics tools and socio-technical system methods can be as complement to each other as supporting more holistic view of the whole problem area. Sometimes it is hard to explain both external and internal factors in a phenomenon. Causal Loop Diagram (CLD) can be useful for the better visualization of relationships of all variables (Senge, 1990).

Step 2: Root Cause Analysis with Critical Success Factors

Case study of Akkermans and Helden (2001) has applied the Causal Loop Diagram (CLD) technique for analyzing vicious and virtuous cycles in ERP implementation. Through a CLD analysis they aimed to test if the Critical Success Factors (CSFs) of ERP implementation according to Somers and Nelson (2001) support an understanding of root causes of success and failures of an ERP implementation. They were also interested in how the CSFs interrelate causally (Akkermans & Helden, 2001). As result of their case study in one organization, they concluded that interdepartmental collaboration and communication reinforce each other (Figure 4 and Figure 5). By using the terms according to the System Dynamics interdepartmental collaboration and communication form “reinforcing loops” (Senge, 1990) of this case study. These are also core processes in the ERP implementation project (Akkermans & Helden, 2001). Without turning to downward spiral by a negative factor, the loop continues reinforcing the cycle and become a so called “virtuous” one (Senge, 1990).

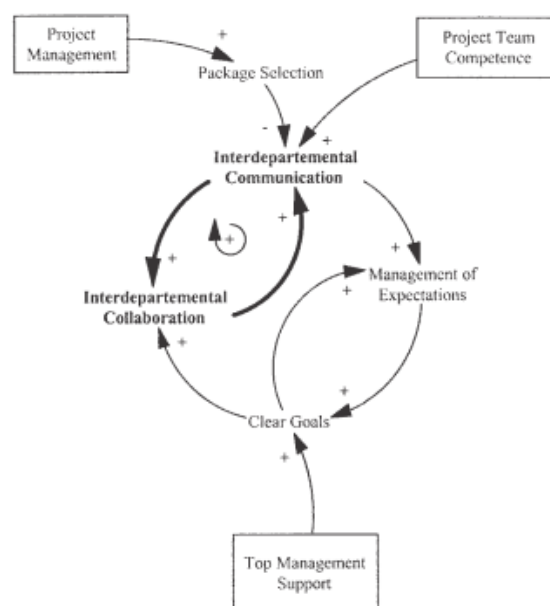


Figure 4: Root causes of the vicious cycle leading to the project crisis (Akkermans & Helden, 2001).

However the so called “vicious cycle” leads to the project crisis in a case of Akkermans and Helden (2001). Figure 4 results from the one-sided communication between the project team and the external consultants which was led by the technical issues. Also a top management was not clear with the goals for the project. Thus, an internal communication which would also need a knowledge sharing in a large range of issues was too weak and biased (Akkermans & Helden, 2001). An ERP provider, who typically has the knowledge of the technical details such as configuration possibilities of the ERP product, has to be able to share this knowledge in the ERP implementing organization. Respectively, an organization who implements the ERP knows the business rules and processes how the ERP system should behave in their specific organizational context (Dittrich, Vaucouleur, & Giff, 2009). Typically, there are not enough people internally to manage all the tasks required within the implementation of ERP (i.e. business process mapping, configuration and project management). As a result, some external experts are hired. There is an earlier research about the consequences if a significant amount of external experts dominate ERP project. Too much power in decision-making of an external organization can harm the project in longer term, especially, if a consultant organization with a dominating knowledge and expertise continues selling more services to their client (Deltour, 2012; Engwall & Jerbrant, 2003).

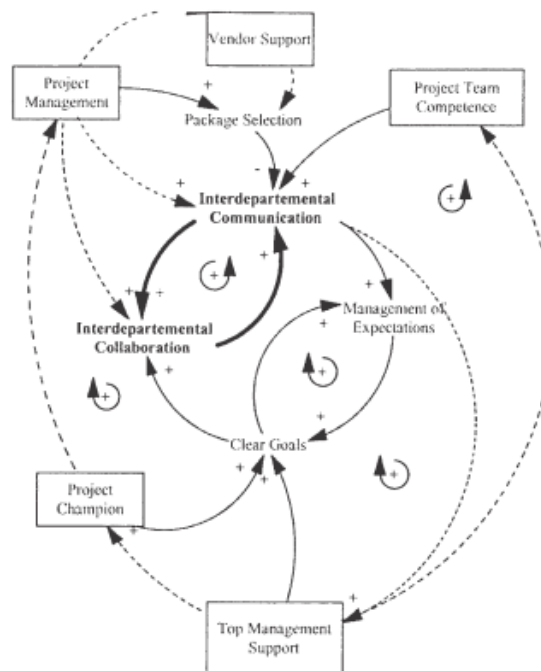


Figure 5: Counter-measures to reverse the vicious cycle into virtuous one (Akkermans & Helden, 2001).

The vicious cycle of the ERP implementation project introduced by Akkermans and Helden (2001) was finally reversed back into the virtuous one with the aid of several right decisions and improvements in the project (Figure 5). Some of the

actions were certainly possible because the researchers were with their good contribution able to replace some of external consultants. This is a benefit of action research type of research in which the researchers can be contributors to the change process in the case organization (Checkland, 1981). Other improvement actions were, for example, assignments of the new people such as a senior manager to take care of the project champion and rearranging the project activities such as knowledge sharing workshops (Akkermans & Helden, 2001). As described in the Figure 5 vendor's stronger support was also one factor reinforcing the communication.

Discussion and Further Development: Understanding of Dynamical ERP Project Nature

It is difficult to predict behaviors of a social system (Forrester, 1971). Thus a project including a specific set of people and environmental elements requires an analysis throughout the project.

Causal Loop Diagram technique can be applied to describe a change process in an ERP implementation. Different factors have either reinforcing or balancing impact on the success of an ERP project in a specific moment (Akkermans & Helden, 2001). However, the case examples in which SD has been used for explaining a dynamic nature of the ERP project are not complete. The causal relations changes in a timeline. Although Critical Success Factors (CSFs) can be realized in inner-organizational ERP implementations, more research is needed in assuming the success in the network of stakeholders across organizational boundaries. For example, an ERP project including many different organizations may have conflicts because of their own values and norms coming from their own culture (Levina & Vaast, 2008).

Complex systems have a nature which cannot be explained with a linear and regular behavior. An ERP system is not necessarily very complex itself. But when the ERP is integrated to the other systems and developed in condition of environmental needs, the complexity increases. Moreover, many ERP implementations in larger size enterprises can concern the organizational units in different countries when local requirements and knowledge management systems should be harmonized for the ERP system (Garud & Kumaraswamy, 2005).

Participatory Design techniques as a part of development activities bring both social and technical aspects to the same design session (Luna-Reyes et al., 2005; Pries-Heje & Dittrich, 2009). This enables a more collaborative and demographic mental model sharing at an early stage for building a shared view. Context specific needs, organizational culture and nature of the system should be taken into account when planning the focus of PD practices. Also more guidelines can be developed for the self-evaluation of the maturity level of organization to collaborate and communicate in an effective manner.

This paper describes a future research proposal how the complex system techniques with a PD approach can be applied to an ERP implementation for making the project dynamical nature more visible among stakeholders. The steps introduced will be tested in a case study within an ERP implementation project.

References

- Ahmad, A., Lyytinen, K., & Newman, M. (2011). The Evolution of Process Models in IS Research: From a Punctuated Social Process Model to a Socio-Technical Process Model. In *ECIS 2011 Proceedings* (Vol. Paper 161). Presented at the European Conference on Information Systems (ECIS). Retrieved from <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1160&context=ecis2011>
- Akkermans, H., & Helden, K. van. (2001). *Vicious and virtuous cycles in ERP implementation: a case study of interrelations between critical success factors*. Eindhoven: [Beta, Research School for Operations Management and Logistics].
- Beyer, H., & Holtzblatt, K. (1998). *Contextual design: defining customer-centered systems*. San Francisco, Calif: Morgan Kaufmann.
- Checkland, P. (1981). *Systems thinking, systems practice*. Chichester [Sussex]; New York: J. Wiley.
- Damsgaard, J., & Karlsbjerg, J. (2010). Seven principles for selecting software packages. *Communications of the ACM*, 53(8), 63. doi:10.1145/1787234.1787252
- Davenport, T. H. (1998). Putting the Enterprise into the Enterprise System. *Harvard Business Review*, (July-August 1998), 11.
- Deltour, F. (2012). ERP Project in SMEs: A Matter of Risks, A Matter of Competencies. A Quantitative Analysis. In *ECIS 2012 Proceedings* (Vol. Paper 138). Presented at the European Conference of Information Systems. Retrieved from <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1137&context=ecis2012>
- DeSanctis, G., & Poole, M. S. (1994). Capturing the Complexity in Advanced Technology Use: Adaptive Structuration Theory. *Organization Science*, 5(2), 121–147. doi:10.1287/orsc.5.2.121
- Engwall, M., & Jerbrant, A. (2003). The resource allocation syndrome: the prime challenge of multi-project management? *International Journal of Project Management*, 21(6), 403–409. doi:10.1016/S0263-7863(02)00113-8
- Forrester, J. F. (1971). Counterintuitive Behavior of Social Systems. *Theory and Decision*, 2, 109–140.
- Garud, R., & Kumaraswamy, A. (2005). Viscious and Virtuous Circles in the Management of Knowledge. *MIS Quarterly*, 29(1), 9–33.
- Giddens, A. (1984). *The constitution of society: outline of the theory of structuration*. Berkeley: University of California Press.

- Kensing, F., & Blomberg, J. (1998). Participatory Design: Issues and Concerns. *Computer Supported Cooperative Work*, 7, 167–185.
- Koh, S. C. L., Gunasekaran, A., & Goodman, T. (2011). Drivers, barriers and critical success factors for ERP II implementation in supply chains: A critical analysis. *The Journal of Strategic Information Systems*, 20(4), 385–402. doi:10.1016/j.jsis.2011.07.001
- Levina, N., & Vaast, E. (2008). Innovating or Doing as Told? Status Differences and Overlapping Boundaries in Offshore Collaboration. *MIS Quarterly*, 32(2), 47.
- Luna-Reyes, L. F., Zhang, J., Ramón Gil-García, J., & Cresswell, A. M. (2005). Information systems development as emergent socio-technical change: a practice approach. *European Journal of Information Systems*, 14(1), 93–105. doi:10.1057/palgrave.ejis.3000524
- Mumford, E. (2000). A Socio-Technical Approach to Systems Design. *Requirements Engineering*, 5(2), 125–133. doi:10.1007/PL00010345
- Orlikowski, W. J. (1992). The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science*, 3(3), 398–427. doi:10.1287/orsc.3.3.398
- Orlikowski, Wanda J. (1996). Improvising organizational transformation over time: A situated change perspective. *Information systems research*, 7(1), 63–92.
- Pawlowski, S., & Raven, A. (2000). Supporting Shared Information Systems: Boundary Objects, Communities, and Brokering. In *ICIS Proceeding 2000* (Vol. 30, p. 11). International Conference on Information Systems. Retrieved from <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1123&context=icis2000>
- Pries-Heje, L., & Dittrich, Y. (2009). ERP Implementation as Design: Looking at participatory design for means to facilitate knowledge integration. *Scandinavian Journal of Information Systems*, 21(2), 33.
- Senge, P. M. (1990). *The fifth discipline: the art and practice of the learning organization*. New York: Doubleday/Currency.
- Somers, T. N., & Nelson, K. (2001). The Impact of Critical Success Factors across the Stages of Enterprise Resource Planning Implementations. In *Proceedings of the 34th Hawaii International Conference on System Sciences - 2001*. Presented at the Hawaii International Conference on System Sciences. Retrieved from <http://www.cs.uu.nl/education/scripties/pdf.php?SID=INF/SCR-2009-065>
- Trist, E. L. (1981). *The evolution of socio-technical systems: a conceptual framework and an action research program*. Toronto: Ontario Ministry of Labour, Ontario Quality of Working Life Centre.
- Wagner, E. L., & Piccoli, G. (2007). Moving beyond user participation to achieve successful IS design. *Communications of the ACM*, 50(12), 51–55. doi:10.1145/1323688.1323694

- Wright, S., & Wright, A. M. (2002). Information System Assurance for Enterprise Resource Planning Systems: Unique Risk Considerations. *Journal of Information Systems*, 16(s-1), 99–113. doi:10.2308/jis.2002.16.s-1.99
- Wu, H., & Cao, L. (2009). Community Collaboration for ERP Implementation. *IEEE Software*, 26(6), 48–55. doi:10.1109/MS.2009.171

Micro-cognition and macro-cognition – Trying to bridge the gap

Hardy Smieszek¹ and Nele Russwinkel²

Keywords: macro-cognitive modelling, modelling methods, research paradigms

Abstract

The term macro-cognition was created by Cacciabue & Hollnagel to distinguish the systemic approach of modelling cognitive systems from the traditional micro-cognitive approach of cognitive science. Especially for cognitive modelling approaches the term macro-cognition is recently becoming more and more popular although the phenomena labelled as macro-cognition are not consistent. Several attempts for connecting micro- and macro-cognitive views are made especially in the cognitive modelling community. This paper describes why macro-cognition was distinguished from micro-cognition and what concerns exist about distinguishing. Furthermore three basic phenomena referred to as macro-cognition identified in the literature will be presented. Two means of connecting micro- and macro-cognition from a cognitive modelling perspective are described in order to try to overcome gaps between micro- and macro-cognitive researches.

Introduction – differentiating between micro- and macro-cognition

Originally Johannssen (1985) introduced a difference between micro- and macro-cognition when he distinguished micro- and macro-levels in an analysis of decision-making situations (Schraagan, Klein & Hoffman, 2008). Later the notion of macro-cognition and the distinction to micro-cognition was revisited by Woods & Roth (1986) in discussing the hierarchy of decision-making situations. Cacciabue & Hollnagel (1995) contrasted macro-cognition with micro-cognition as they were arguing for a new view of human-machine systems design different from the information-processing approach (Schraagan, Klein & Hoffman, 2008): In their opinion cognitive functions should be studied in natural settings (on the job, in realistic tasks – macro-cognitive) rather than in controlled, isolated and sometimes artificial laboratory settings (micro-cognitive) (Klein, Ross, Moon, Klein, Hoffman & Hollnagel, 2003).

This applies for cognitive modelling approaches as well. Cacciabue & Hollnagel (1995) argue for a more systemic view of models of sociotechnical systems in the tradition of cognitive systems engineering (Hollnagel & Woods, 1983) rather than models of single cognitive functions in isolation of the environment the individual

¹ Technische Universität Berlin, research group prometei, Marchstrasse 23, 10587 Berlin

² Technische Universität Berlin, Department of Cognitive Modeling in dynamic Human-Machine Systems, Marchstrasse 23, 10587 Berlin

performs its tasks in. Therefore the emphasis should be more on relating the model more closely to the applied problem than to underlying cognitive functions of the human alone.

Macro-cognition – trying to define the concept

Schraagan, Klein & Hoffman (2008) define macro-cognition as: “the study of cognitive phenomena found in natural settings, especially (but not limited to) cognitive work conducted in complex sociotechnical contexts” (p.8).

The interest of macro-cognitive research and modelling is not on the theoretical accounts of how cognition takes place in the human mind, but on the overall performance of the human-machine system as a whole (called joint cognitive system; Cacciabue & Hollnagel, 1995; Hollnagel & Woods, 2005). “The interest is more on how tasks are performed and how well they serve to achieve their goals than on the details of what goes on in the mind while they are performed.” (Cacciabue & Hollnagel, 1995, p.57). The approach is therefore functional rather than structural.

Practitioners in natural settings rarely focus on micro-cognitive processes. Instead, they are concerned with macro-cognitive phenomena like functions and processes (Klein, Ross, Moon, Klein, Hoffman & Hollnagel, 2003). Key macro-cognitive functions are therefore defined by Klein et al. (2003): decision making, sensemaking, planning, adaption/replanning, problem detection and coordination. Furthermore, supporting processes are defined as means to achieve the primary macro-cognitive functions: maintaining common ground, developing mental models, mental simulation & storybuilding, managing uncertainty & risk, identifying leverage points, managing attention. These functions and processes are subject of macro-cognitive research.

“Mind the gap” – concerns about differentiating

There is not only agreement but also some serious concerns about differentiating micro- and macro- cognition. Staszewski (2008) for example warns against seeing macro- and micro- cognition as competing or antagonist frameworks (Schraagan, Klein & Hoffman, 2008). Also Flach (2008) argues that with macro-cognition “*we have invented yet another name for a ‘bird’ that we still know precious little about.*” (p.28). It is not clear if the different constructs represent distinct phenomena or simply different perspectives on a single phenomenon. One obvious gap that Flach (2008) identified is between the view of cognition that shapes experimental paradigms and laboratory tasks and the view of cognition derived from more ethological studies of human performance. This is one of the features used to distinguish macro-cognition from “traditional” psychological research. The distinction between micro- and macro- cognition rises from the concern that generalizing from classical laboratory research on human performance to the problem of natural human experiences is inadequate. It is therefore not a different kind of human cogni-

tion that is examined but the paradigm of research is different: “*All Cognition is macro. It is only the paradigms of science that are micro.*” (Flach, 2008, p.16).

That is why it is argued here and it should always be clearly stated, that micro- and macro-approaches of science are not competing or opponent but should rather be used complementary to each other. Currently there is serious effort connecting micro- and macro-cognitive approaches not only in experimental psychology but especially in the cognitive modelling community. Attempting to overcome the gap, several researchers proposed their own definition of macro-cognition. However, these definitions represent slightly different views of what macro-cognition is. Nevertheless, these slightly different views of what macro-cognition is are not always clearly defined. They at large refer to the same roots, but mean slightly different things and are therefore running the risk of confusion. Three such views of macro-cognition have been identified in the literature so far, and will be presented in the next paragraph.

Three views of macro-cognition

Three different meanings of the term macro cognition have been identified so far, which are present within the literature today attempting to overcome the gap between micro- and macro-cognition. The first two meanings stem from the community of modelling cognition and performance of human operators, the third from research concerning teams.

The first meaning is described by Smieszek, Manske, Hasselberg, Russwinkel & Moehlenbrink (2013) and follows the tradition of macro-cognitive modelling proposed by Cacciabue & Hollnagel (1995). Hereby macro-cognition is seen as the performance of experts in realistic tasks, where the emphasis of the model is on the overall impact of cognitive components on the modelled system. It is argued that especially for modelling sociotechnical systems it is not only necessary to model human cognition but also the system to be controlled in order to generate insights into the performance of the joint cognitive system. It can be denoted as cognition in the wild (Hutchins, 1996) or cognitive simulation (Cacciabue, 1998).

The second meaning can be referred to as “macro cognition in the mind” or the “macro architecture hypothesis”. As proposed by West, Hancock, Somers, MacDougall & Jeanson (2013) a macro cognitive architecture exists in the brains of individuals and enables humans to apply information processing abilities (micro-cognition) to complex, dynamic, multi-agent, real world tasks (macro-cognition). This means that micro-cognitive functions serve as a means for creating macro-cognitive functions in a way that neuronal functions serve as a means in creating micro cognitive functions (see West, Hancock, Somers, MacDougall & Jeanson, 2013). The knowledge about macro-cognitive phenomena and coherences is therefore located in the human mind.

The third meaning can be referred to as “team cognition”, “shared cognition” or “shared mental models”. It is described by Grauel, Kluge & Adolph (2013), fol-

lowing Letsky, Warner, Fiore, & Smith (2008) and describes cognitive processes that arise in teams of several human operators working together in sociotechnical contexts. Especially the process of transforming internalised team knowledge (knowledge of the individual) into externalised team knowledge (knowledge of the team) through processes of individual knowledge acquisition and knowledge acquisition in teams is seen as macro cognition. The three faces are depicted graphically in figure 1.

Connecting micro- and macro cognition – trying to bridge the gap

It is already realized that micro- and macro-cognition are not distinct or antagonist concepts but instead two sides of the same coin, two ways of seeing the world, two paradigms of research. Therefore it is possible and also necessary to overcome obvious gaps and try to connect macro- and micro-research, as they can obviously learn from and inspire each other. The three meanings of macro-cognition explained earlier are examples of attempts to overcome these gaps.

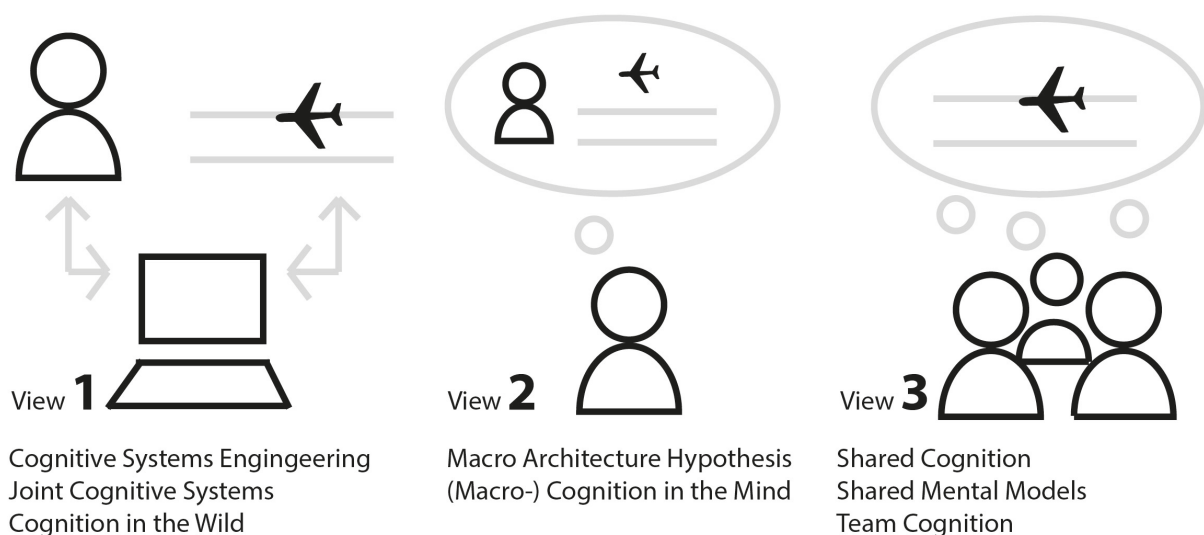


Figure 1: The three views of macro-cognition.

Especially in the context of cognitive modelling there are two possibilities which can representatively be compared to the approaches of Smieszek, Manske, Hasselberg, Russwinkel & Moehlenbrink (2013) and West, Hancock, Somers, MacDougall & Jeanson (2013). The first approach is trying to integrate micro cognitive functions into a macro cognitive model of the airport traffic control system (called MATriCS) to investigate the impact of cognitive components on the performance of the cognitive system as a whole (e.g. human and machine). By *downscaling* the model of cognitive processes a higher level of granularity can be reached, as macro cognitive modelling traditionally does not claim to model every single cognitive phenomenon in the very detail micro cognitive modelling does. This is done e.g. by implementing a working memory model into the model of the

air traffic controller. Further cognitive functions like attention switching, perception of stimuli or decision making processes are modelled, which serve for the controller model in conducting the air traffic control task.

Several downscaling processes of the model are still in progress e.g. for the decision making process which initially was modelled with a low granularity but is now further refined to take prospective decision making and mental simulation into account (see Smieszek & Joeres, this issue). Some further gaps could be overcome in this modelling process, as several findings of micro-cognitive research were already used in the MATriCs model (e.g. Miller's 1956 claim of a working memory capacity of 7 ± 2 chunks). However, further need of basic research by using micro-cognitive approaches was implicated (e.g. the question of memory updating frequency in mental simulation tasks, Smieszek & Joeres, 2013). This shows how both research paradigms can inspire and support each other.

Another approach of overcoming gaps is proposed by West, Hancock, Somers, MacDougall & Jeanson (2013) by *upscaling* from a micro-cognitive architecture in order to model macro-cognitive functions using the macro architecture hypothesis. Even though it is argued that architectures built to model micro-cognitive tasks cannot scale up to usefully model macro-cognitive behaviour, West & Nagy (2007) counter, however, that they can. They further hypothesise that the difference between micro- and macro-cognition is analogous to the difference between neural models and micro-cognitive models - there is a macro level architecture that is built on a micro level architecture (West & Pronovost, 2009). West et al. (2013) showed that by using the micro-cognitive architecture ACT-R (Anderson & Lebiere, 1998) in combination with the SGOMS (Sociotechnical GOMS) framework the macro-cognitive functions of mediation (concerning strategies for mediating between different opinions across multiple agents) and team play (especially reactions to interruptions and unexpected events) could be modelled.

Summary and Discussion

The terms micro- and macro-cognition are often discussed especially in the naturalistic decision making community (e.g. Klein, 2008). It was introduced to distinguish the naturalistic or ecological approach of psychology from the traditional laboratory and information processing approaches. Nevertheless, it is argued that micro- and macro-cognition are not distinct phenomena but two sides of the same coin and that obvious gap should be overcome. Two approaches of bridging gaps between micro- and macro-cognitive researches especially in the field of cognitive modelling were presented. Both try to introduce a more holistic kind of modelling including micro- as well as macro-cognitive functions: One by downscaling and implementing micro-cognitive functions into a macro-cognitive model. The other by upscaling from a micro-cognitive architecture to macro-cognitive functions (see also McAndrew, Banks & Gore, 2008). As explained before, it is not only necessary but also very useful to overcome obvious gaps and to connect both research

traditions. Indeed some questions about human cognition can only be answered using basic research and a micro-cognitive approach and the macro-cognitive approach is needed to apply these findings to real world problems.

Literature

- Anderson, J. R. & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cacciabue, P. C. (1998). *Modelling and Simulation of Human Behaviour in System Control*. London: Springer.
- Cacciabue, P. C., & Hollnagel, E. (1995). Simulation of Cognition: Applications. In J.-M. Hoc, P. C. Cacciabue, & E. Hollnagel (Eds.), *Expertise. Expertise and technology. Cognition & human-computer cooperation* (pp. 55–73). Hillsdale, N.J: Lawrence Erlbaum Associates.
- Flach, J. M. (2008). Mind the Gap: A skeptical view of Macrocognition. In J. M. Schraagan, L. G. Militello, T. Ormerod & R. Lipshitz (eds.), *Naturalistic Decision Making and Macrocognition* (pp. 27-40). Aldershot: Ashgate.
- Grael, B. M., Kluge, A., & Adolph, L. (2013). Analyse vorausgehender Bedingungen für die Unterstützung makrokognitiver Prozesse in Teams in der industriellen Instandhaltung. In: D. Soeffker & A. Kluge (eds.). *Kognitive Systeme*, Universität Duisburg-Essen.
- Hollnagel, E. & Woods, D. D. (1983). Cognitive Systems Engineering: New wine in new bottles. *International Journal of Man-Machine-Studies*, 18, 583-600.
- Hollnagel, E. & Woods, D. D. (2005). *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. Boca Ranton: Taylor & Francis.
- Hutchins, E. (1996). *Cognition in the Wild*. MIT press.
- Johanssen, G. (1985). Architecture of man-machine decision making systems. In E. Hollnagel, G. Mancini, and D.D. Woods (eds.), *Intelligent Decision Support in Process Environments* (pp. 327-339). Berlin: Springer.
- Klein, G. (2008). Naturalistic Decision Making. *Human Factors*, 50(3), 456-460.
- Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffinan, R. R., and Hollnagel, E. (2003). Macrocognition. *IEEE Intelligent Systems*, 81-85.
- McAndrew, C., Banks, A., & Gore, J. (2008). Bridging Macrocognitive/ Microcognitive Methods: ACT-R under Review. In J. M. Schraagan, L. G. Militello, T. Ormerod & R. Lipshitz (eds.), *Naturalistic Decision Making and Macrocognition* (pp. 317-350). Aldershot: Ashgate.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Schraagan, J. M., Klein, G. & Hoffman, R. R. (2008). The macrocognition framework of naturalistic decision making. In J. M. Schraagan, L. G. Militello, T.

- Ormerod & R. Lipshitz (eds.), *Naturalistic Decision Making and Macro-cognition* (pp. 3-25). Aldershot: Ashgate.
- Smieszek, H., Manske, P., Hasselberg, A., Rußwinkel, N. & Möhlenbrink, C. (2013). Cognitive Simulation of limited Working Memory Capacity applied to an Air Traffic Control Task. In: R. West & T. Stewart (eds.), *Proceedings of the 12th International Conference on Cognitive Modeling*, Ottawa: Carleton University. [Available as open source download]: <http://iccm-conference.org/2013-proceedings/170/0170-paper.pdf> [07.08.2013]
- Staszewski, J. J. (2008). Cognitive Engineering Based on Expert Skill: Notes on Success and Surprises. In J. M. Schraagan, L. G. Militello, T. Ormerod & R. Lipshitz (eds.), *Naturalistic Decision Making and Macro-cognition* (pp. 317-350). Aldershot: Ashgate.
- West, R. L., & Nagy, G. (2007). Using GOMS for Modelling Routine Tasks Within Complex Sociotechnical Systems: Connecting Macro-cognitive Models to Micro-cognition. *Journal of Cognitive Engineering and Decision Making*, 1(2), 186-211.
- West, R. L. & Pronovost, S. (2009). Modeling SGOMS in ACT-R: Linking Macro- and Micro-cognition. *Journal of Cognitive Engineering and Decision Making*, 3(2), 194-207.
- West, R. L., Hancock, E., Somers, S., MacDougall, K. & Jeanson, F. (2013). The Macro Architecture Hypothesis: Applications to Modelling Teamwork, Conflict resolution, and Literary Analysis. In: R. West & T. Stewart (eds.), *Proceedings of the 12th International Conference on Cognitive Modeling*, Ottawa: Carleton University. [Available as open source download]: <http://iccm-conference.org/2013-proceedings/197/0197-paper.pdf> [07.08.2013]

Face Interface: Using voluntary gaze direction and facial muscle activations for human-computer interaction

Outi Tuisku¹, Veikko Surakka⁸⁹, Ville Rantanen² and Jukka Lekkala⁹⁰

Keywords: eye tracking, facial muscle activation, multimodality, prototype, human-computer interaction

Abstract

The aim of the research on Face Interface has been to iteratively develop and study a new Face Interface prototype for human-computer interaction. It combines the use of two modalities: voluntarily controlled gaze direction and facial muscle activations for pointing and selecting objects, respectively, in a graphical user interface. The measurement technologies were embedded in a wearable, eyeglass-like frame that houses both: eye tracker and capacitive sensor(s) to measure the levels of facial activations.

The Face Interface technology has been used in a series of studies as the pointing and selection device, as follows:

- The technical functionality of Face Interface has been rigorously tested as such. Especially the speed and accuracy has been tested in a series of experimental studies using simple point-and-select tasks.
- Face Interface has been used in on-screen text entry tasks. For that, three on-screen keyboard layouts were designed and tested. Then, the most promising of them was experimentally tested for typing.

Face Interface

Face Interface has been a collaborative project between University of Tampere (UTA) and Tampere University of Technology (TUT) that was funded by the Academy of Finland in 2007-2010. The project resulted in a concept named after it, that is, Face Interface. Professor Surakka (UTA) has been the responsible investigator of the whole project. Four research groups have been working in the project. Professor Lekkala's (TUT) research group was in charge of the technological studies (i.e. building the device), and Professor Juhola's (UTA), Professor Hyttinen's (TUT), and Professor Surakka's research groups were in charge of the software development. Finally, Professor Surakka's group has been responsible of the empirical testing of the prototype.

¹ Tampere Unit for Computer-Human Interaction, School of Information Sciences, University of Tampere, Finland

² Department of Automation Science and Engineering, Tampere University of Technology, Finland

Face Interface is a wearable device that houses both: an eye tracker to measure the gaze direction and capacitive sensor(s) to detect facial muscle activations. It is built on the frames of protective glasses, and thus, is worn like eyeglasses. It consists of video camera(s), one for imaging the eye (i.e. eye camera) and the other for imaging the surrounding world (i.e. scene camera). Face Interface is used so that the objects are pointed by gazing at them and then selected by voluntary facial muscle activations (e.g. by frowning).

In the course of the thesis work, three different Face Interface versions have been used. The first one (see Figure 1) was wired and had only one camera. It used a commercial low-cost Bluetooth web camera (i.e. Creative Live! Cam Notebook) as an eye camera which had a frame rate of 30 frames per second. It had one capacitive sensor placed in the bridge of the nose of the prototype so that it was able to monitor only the frowning related movements. Also, a chin rest was used in experimental testing to prevent any involuntary head movements.

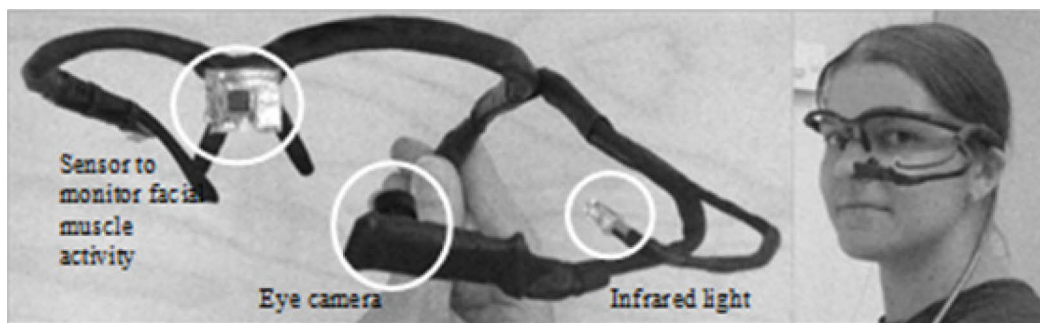


Figure 1. On the left the prototype and on the right person wearing it (figure adapted from Tuisku et al., 2011).

The second prototype version (see Figure 2) was made wireless and it was able to monitor either frowning or raising the eyebrows related movements, depending on a preference of a participant (Rantanen et al., 2011; Tuisku et al., 2012). In addition, a scene camera was added to compensate head movements. It was used so that it imaged the computer display and identified the six physical markers that were attached to the display. Using this data, the head movements could be compensated.



Figure 2. On the left the wireless prototype device and on the right person wearing it.

The third prototype version (see Figure 3) was further improved by using five capacitive sensors to detect three different facial activations: frowning, raising the eyebrows, and smiling (Rantanen et al., 2012). This was achieved by placing the capacitive sensors above both eyebrows, in front of the forehead, and in front of both cheeks. Scene camera image processing was improved by detecting the screen bezel to avoid attaching physical markers to the display.



Figure 3. On the left the box for wireless operations and on the right person wearing the prototype.

Methods

The Face Interface prototype has mainly been studied using simple pointing and selection tasks (Rantanen et al., 2011, 2012; Tuisku et al., 2011, 2012, *submitted*). The pointing tasks proceeded as follows. First, the participant was to point a home square by gaze (i.e. by looking at it) and then select it with a voluntary facial activation. Then, the participant was to point and select a target circle similarly. When the participant's gaze was inside an object, the object was highlighted as an indication that the selection could be made. When the object(s) were selected, they disappeared and reappeared again in different position in the display after a pause of 2000 ms. The diameter of the home square was kept constant and it was 30 mm. Three different diameters were used for the target circle and they were 25, 30, and 40 mm. The target circle always appeared in one of the eight angles (i.e. two verti-

cal, two horizontal, and four diagonal) in relation to the home square. The pointing distance was measured from the center point of the home square to the center point of the target circle. Different pointing distances were used in different studies and were, for example, 60, 120, 180, and 240 mm.

The pointing task time was measured from the selection of the home square to the selection of the target circle. If the first click of the target circle was outside the object, that trial was marked as an erroneous one and was excluded from the pointing task time analysis.

In addition to the pointing task, subjective ratings have also been collected in every study. The used scales were first introduced by Surakka et al. (2004) and were: general evaluation (i.e. from bad to good), difficulty (i.e. from difficult to easy), speed (i.e. from slow to fast), accuracy (i.e. from inaccurate to accurate), enjoyableness (i.e. from unpleasant to pleasant), and efficiency (i.e. from inefficient to efficient). The scales varied from -4 (e.g. bad experience) to +4 (e.g. good experience), and 0 represented a neutral experience (e.g., not slow nor fast).

In the final experiments, Face Interface was used for entering text (Tuisku et al., 2013). First, three different on-screen keyboard layouts were designed and tested. To compare them against each other, a feature that was used was randomization. That is, the places of letters were randomized again after every typed word. This approach was chosen to rule out the familiarity with traditional QWERTY layout which could have resulted in the case that comparison between the Face Interface and the mouse would distorted the results to prefer the mouse. After the most functional layout was identified, entering text with Face Interface was compared to entering text with a regular computer mouse.

Results and Discussion

The results have shown that Face Interface functions promisingly as a pointing and selection technique. Along the iterations, significant improvements have been achieved in the pointing task times (i.e. from 2.5 seconds (Tuisku et al., 2011) to 1.3 seconds (Tuisku et al., *submitted*)). Also the error rates have been significantly improved throughout this thesis work from 28.5% (Tuisku et al., 2011) to 16% (Tuisku et al., *submitted*).

In the course of this thesis work, the area that the Face Interface is the most functional for pointing and selecting purposes was revealed (Tuisku et al., 2012). The area covers pointing distances from 60 mm to 260 mm. This means that in the middle of the widescreen display it is easier to point and select targets than it is at the edges of the display.

Interestingly, the subjective ratings were on the higher level in the first study (Tuisku et al., 2011) than in the other studies (Tuisku et al., 2012, *submitted*) even though the pointing task times show the contrary.

Text entry studies showed that the Face Interface can be used also for this purpose (Tuisku et al., 2013). The text entry rates for the first time Face Interface users have been encouraging, that is, 4 words per minute (wpm) on average. As compared to computer mouse, Face Interface was slower. This is not surprising considering that all our participants in each experiment had been using the mouse for several years. We note that the Face Interface technology has been shown to be a very easy to learn. It takes basically less than 5 minutes to learn the use of the technology. In comparison to many other alternative new interaction methods this is an advantage. It also suggests that face-based human-technology interaction methods can be competitive future technologies for the reason that they rely on behaviors that humans use naturally for human-human interaction.

Acknowledgements

This research was funded by the Academy of Finland (project numbers 115997 and 116913) and the Finnish Doctoral Programme in User-Centered Information Technology (UCIT).

Literature

- Rantanen, V., Vanhala, T., Tuisku, O., Niemenlehto, P.-H., Verho, J., Surakka, V., Juhola, M., & Lekkala, J. (2011). A wearable, wireless gaze tracker with integrated selection command source for human-computer interaction. *IEEE Transactions on Information Technology in BioMedicine*, 15(5), 795-801. doi:10.1109/TITB.2011.2158321
- Rantanen, V., Verho, J., Lekkala, J., Tuisku, O., Surakka, V., & Vanhala, T. (2012). The effect of clicking by smiling on the accuracy of head-mounted gaze tracking. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, 345-348. doi:10.1145/2168556.2168633
- Surakka, V., Illi, M., & Isokoski, P. (2004). Gazing and frowning as a new human-computer interaction technique. *ACM Transactions on Applied Perception*, 1(1), 40-56. doi: 10.1145/1008722.1008726
- Tuisku, O., Rantanen, V., Špakov, O., Surakka, V., & Lekkala, J. (Submitted). Pointing and selecting with facial activity. *Submitted*.
- Tuisku, O., Surakka, V., Gizatdinova, Y., Vanhala, T., Rantanen, V., Verho, J., & Lekkala, J. (2011). Gazing and frowning to computers can be enjoyable. In *Proceedings of the Third International Conference on Knowledge and Systems Engineering*, 211-218. doi:10.1109/KSE.2011.41
- Tuisku, O., Surakka, V., Vanhala, T., & Lekkala, J. (2012). Text entry by gazing and smiling. *Advances in Human-Computer Interaction*, Article ID 218084, 13 pages. doi: 10.1155/2013/218084
- Tuisku, O., Surakka, V., Vanhala, T., Rantanen, V., & Lekkala, J. (2012). Wireless face interface: Using voluntary gaze direction and facial muscle activations

for human-computer interaction. *Interacting with Computers*, 24(1), 1-9.
doi:10.1016/j.intcom.2011.10.002

Face-hugging device for technology mediated human-human interaction

Hanna Venesvirta¹, Veikko Surakka⁹¹, Jani Lylykangas⁹¹, Ville Rantanen², Oleg Špakov⁹¹, Jarmo Verho⁹², Akos Vetek³ and Jukka Lekkala⁹²

Keywords: Nonverbal communication, facial expressions, capacitive measurement, technology mediated human-human interaction

Abstract

Nonverbal gestures, specifically extracted from the human head area, are proven to be an important source of supporting information for human communication. However, there is not that many means to measure and convey this socially important information for technology mediated human-human interaction. We have developed a novel prototype to detect nonverbal behaviour from the human head area. With controlled experiments we have investigated the detection of nonverbal expressions with a capacitive detection method. Moreover, we have run an experiment to compare the use of nonverbal gestures produced with either the prototype or a mouse. We have found out that the prototype can be used to detect both different facial expressions and the different intensity levels of expressions. Further, we have found out, for instance, that the prototype was preferred over the mouse for producing nonverbal gestures. These results encourage us to develop further this new alternative for technology mediated nonverbal communication.

Introduction

Imagine that you are having a conversation with your friend via a social media service or a virtual environment. While you are chatting, you possibly would like to underline that something has been pleasant with a smile. You might also have a need to show that you understand or agree on something your friend brings out. Would it not be exciting that you could do this automatically by smiling or nodding while reading your friend's writing? In turn it would be nice as well to know how your friend responds to your opinions. What kind of opportunities do you have today to express yourself if you are pleased or confused when you chat? Besides writing your thoughts to text you cannot do more than use emoticons or, in more advanced services like Second Life, animated gestures, which in turn you need to type with a keyboard or select by using a mouse. Emoticons or animations

¹ Research Group for Emotions, Sociality, and Computing, Tampere Unit Computer-Human Interaction, School of Information Sciences, University of Tampere, Tampere, Finland

² Sensor Technology and Biomeasurements, Department of Automation Science and Engineering, Tampere University of Technology, Tampere, Finland

³ Media Technologies Laboratory, Nokia Research Center, Espoo, Finland

might not be fully in line with the message you would like to express. Thus, it would be great to use some novel, futuristic device, which could easily convey your facial behaviour and head gestures you make in real time so that your friend could understand your nonverbal actions.

We have been developing a new prototype for measuring nonverbal behaviour from the human head area. It is an easy-to-wear, light-weight device (see Figure 1) which can measure head movements and changes in facial expressions. The measured data is sent wirelessly to a computational device which interprets the data with signal processing algorithms. After this, the data can be used, for example, for emotion related (e.g., smiling and frowning) communication in virtual environments like Second Life.

The goal of the research presented here is to investigate how this novel prototype system can be used to measure and convey nonverbal behaviour from the human head area during human-technology interaction or technology mediated human-human interaction (see Figure 1). To achieve the goal, the research will consist of three closely connected subtasks:

- Experimental development of the prototype as a measurement device: the aim is to investigate, which nonverbal gestures can be most optimally measured with the device. Further, the development of signal processing methods is required as the measured gestures should be distinguished from each other.
- The design of suitable interpretations on the basis of the input signals: the aim is to investigate, how the measured signals of nonverbal behaviour can be conveyed to another person. This is an important step of the research as the prototype device provides the possibility of detecting nonverbal gestures, which need to be conveyed with sufficient clarity to another person.
- Investigation of using the prototype device as a communication and interaction device: the aim is to investigate, how the prototype can be used as a supplemental device to express nonverbal behaviour when people are, for example, using technology mediated communication for human to human interaction.



Figure 1: Face-hugging prototype device.

Background

It is known that human communication relies heavily on nonverbal behaviour. Facial expressions and other information (e.g., head movements like nodding and shaking) extracted from the human head area are a rich source of social and emotional interaction (e.g., Surakka & Hietanen, 1998; Afifi, 2006). Both spontaneous and voluntary expressions and gestures are often used in human to human interaction (Levenson *et al.*, 1990; Surakka & Hietanen, 1998). In addition, the information measured from the head area is proven to be a useful modality for human-technology interaction, as facial expressions can be activated both spontaneously and voluntarily. Previously, voluntary facial expressions have been used as a part of multimodal pointing and selecting techniques together with gaze tracking (e.g., Surakka *et al.*, 2004; 2005; Tuisku *et al.*, 2011; 2012; 2013).

At the same time when both spontaneous and voluntary nonverbal signals are important in direct human-human interaction, they have been proven to be important also in technology mediated human-human interaction. In virtual environments different kinds of nonverbal behaviours, like the gaze direction, gestures, and facial expressions can be used to make the interaction situation more natural and give important information about user's social presence, among other things. For example, nonverbal cues for turn taking in virtual environments have been found to be useful for communication situations. It has also been found that even very simple nonverbal gestures, like waving hello, were spontaneously and often used in virtual environment conversations (e.g., Allmendinger, 2010). Virtual avatars can be both pre-animated and animated on real-time to convey nonverbal signals. For example, for real-time animation user's facial expressions and head movements can be recorded and identified by using computer vision (e.g., Budford & Blake, 1999; Kapoor & Picard, 2001; Bailenson *et al.*, 2006). It is also possible to control virtual agents with facial muscle activations, measured with electromyography (EMG) (e.g., Vanhala *et al.*, 2012). Using computer vision techniques for expression recognition might require markers on specific locations on the user's

face. Suitable lighting is required, as well as camera(s). EMG measurement requires electrodes attached on accurate locations on the user's skin and wired to the measuring device.

The novel prototype device shown in Figure 1 offers a new alternative for detecting nonverbal signals for technology mediated interaction. This alternative device can solve many of the challenges mentioned above. The device has been developed in collaboration with University of Tampere, Tampere University of Technology, and Nokia Research Center. The prototype is a wireless, easy-to-wear, headphone-based facial behaviour and head movement measurement device. Movement monitoring through 3D accelerometer and 3D gyroscope can be used for measuring and monitoring, for example, head movements.

For the non-invasive detection of facial expressions, we have developed a novel, contact-free method, which relies on proximity detection between capacitive sensors and the facial skin. The method was presented by Rantanen *et al.* (2010), and the method has been proven to be functional as a part of a multimodal pointing device (e.g., Tuisku *et al.*, 2011; 2012; 2013), for detecting different kinds of voluntarily generated facial expressions (Rantanen *et al.*, 2012), and their intensity levels (Rantanen *et al.*, in press). The targeted use of the face-hugging device in present work is to enable the detection of, for instance, smiling, grimacing, frowning, and lifting eyebrows with the capacitive sensing method, together with different kinds of head movements, detected with the inertial sensors. The measurement data is sent wirelessly to a computational device. The device which receives the data interprets it with developed signal processing algorithms, and after this the data can be used, for example, for communicating nonverbal behaviour.

Status of the Work, Results, and Future Work

We have conducted a controlled experiment to test the functionality of the capacitive method to detect different facial gestures. As a result we can conclude that different facial movement types can be located relatively well to the correct facial areas (Rantanen *et al.*, 2012). Further, we have conducted a controlled experiment to investigate the possibilities of the capacitive detection method to distinguish different intensity levels of varying expressions. The different intensity levels were verified with concurrent EMG measurements. The results showed that medium and high intensity facial activations could be reliably detected (Rantanen *et al.*, in press).

In addition, we have conducted a comparative study, in which we compared the interaction with the novel prototype and a regular computer mouse. The aim was to study, for instance, how fast the nonverbal interaction was with the prototype compared to the mouse and which interaction method the participants preferred to use for producing nonverbal gestures. The prototype turned out to be significantly faster than the mouse, and in addition the participants preferred using it. The partici-

pants rated that the prototype was more pleasant, faster, and more enjoyable to use than the mouse, among other things.

The results of our experiments encourage us to further develop this technology in such a way that it allows the transmission of the information related to, for instance, emotional behaviour. In the future, we will develop further signal processing methods to distinguish different facial behaviours. Also, we will design dynamic visualisations that reflect and represent the signals measured with the prototype. The representations can then be used for technology mediated nonverbal communication. For this purpose we will develop methods to interpret and visualise the measured signals from expressions to convey the nonverbal behaviour to other persons. One promising alternative is to use simplistic biological motion models, in which the idea of visual perception is based on movement patterns which are typical for certain human activities like walking, running, or dancing (Johansson, 1973). The interpretations of different kinds of visualisations will be investigated experimentally. Further, we will test the prototype device in a communication scenario between two people with the help of the prototype and the previously mentioned developments of interpretations.

Acknowledgements

This research has been funded and supported by the Finnish Doctoral Program in User-Centered Information Technology (UCIT), Nokia Research Center, Finnish Funding Agency for Technology and Innovation (Tekes), Finnish Cultural Foundation, Nokia Foundation, and International Measurement Confederation (IMEKO).

References

- Allmendinger, K. (2010). Social Presence in Synchronous Virtual Learning Situations: The Role of Nonverbal Signals Displayed by Avatars. *Educational Psychology Review* 22(1), 41-56. doi: 10.1007/s10648-010-9117-8
- Afifi, W. A. (2006). Nonverbal communication. In B. Whaley & W. Samter (Eds.), *Explaining communication: Contemporary theories and exemplars* (pp. 39-60). Mahwah, NJ: Lawrence Erlbaum
- Bailenson, J., Yee, N., Merget, D. & Schroeder, R. (2006). The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction. *Presence: Teleoperators and Virtual Environments* 15(4), 359-372. doi:10.1162/pres.15.4.359
- Burford D. & Blake, E. (1999). Real-time facial animation for avatars in collaborative virtual environments. *South African Telecommunications Networks and Applications Conference '99*, 178-183.

- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics* 14(2), 201-211. doi: 10.3758/BF03212378
- Kapoor, A. & Picard, R. W. (2001). A real-time head nod and shake detector. *Proceedings of the 2001 workshop on Perceptive user interfaces*, 1-5. doi: 10.1145/971478.971509
- Levenson, R. W., Ekman, P. & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology* 27(4), 363-384. doi:10.1111/j.1469-8986.1990.tb02330.x
- Pinheiro, C. G., Naves, E. L. M., Pino, P., Losson, E., Andrade, A. O. & Bourhis, G. (2011). Alternative Communication Systems for people with severe motor disabilities: a Survey. *BioMedical Engineering OnLine* 10(1), 1-28. doi:10.1186/1475-925X-10-31
- Rantanen, V., Kumpulainen, P., Venesvirta, H., Verho, J., Špakov, O., Lylykangas, J., Vetek, A., Surakka, V. & Lekkala, J. (2012). Capacitive facial activity measurement. *Proceedings of the XX Imeko World Congress*, Busan, South Korea, September 2012.
- Rantanen, V., Niemenlehto, P.-H., Verho, J. & Lekkala, J. (2010). Capacitive facial movement detection for human-computer interaction to click by frowning and lifting eyebrows. *Medical and Biological Engineering and Computing* 48(1), 39-47. doi: 10.1007/s11517-009-0565-6
- Rantanen, V., Venesvirta, H., Špakov, O., Verho, J., Vetek, A., Surakka, V., & Lekkala, J. (in press). Capacitive Measurement of Facial Activity Intensity. To appear in *IEEE Sensors Journal*. Retrieved from <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6542641>
- Surakka, V. & Hietanen, J. K. (1998). Facial and emotional reactions to Duchenne and non-Duchenne smiles. *International Journal of Psychophysiology* 29(1), 23-33. doi: 10.1016/S0167-8760(97)00088-3
- Surakka, V., Illi, M. & Isokoski, P. (2004). Gazing and frowning as a new human-computer interaction technique. *ACM Transactions on Applied Perception* 1(1), 40-56. doi: 10.1145/1008722.1008726
- Surakka, V., Isokoski, P., Illi, M. & Salminen, K. (2005). Is it better to gaze and frown or gaze and smile when controlling user interfaces? *Proceedings of the 11th International Conference on Human-Computer Interaction (HCI'05)*, CD-ROM, 7 pages, July 2005.
- Tuisku, O., Surakka, V., Gizatdinova, Y., Vanhala, T., Rantanen, V., Verho, J. & Lekkala, J. (2011). Gazing and Frowning to Computers Can Be Enjoyable. *Third International Conference on Knowledge and Systems Engineering*, 211-218. doi: 10.1109/KSE.2011.41

- Tuisku, O., Surakka, V., Rantanen, V., Vanhala, T. & Lekkala, J. (2013). Text Entry by Gazing and Smiling. *Advances in Human-Computer Interaction*, vol. 2013, Article ID 218084, 13 pages. doi: 10.1155/2013/218084
- Tuisku, O., Surakka, V., Vanhala, T., Rantanen, V. & Lekkala, J. (2012). Wireless Face Interface: Using voluntary gaze direction and facial muscle activations for human–computer interaction. *Interacting with Computers* 24(1), 1-9. doi: 10.1016/j.intcom.2011.10.002
- Vanhala, T., Surakka, V., Courgeon, M. & Martin, J.-C. (2012). Voluntary facial activations regulate physiological arousal and subjective experiences during virtual social stimulation. *ACM Transactions on Applied Perception* 9(1), 1-21. doi: 10.1145/2134203.2134204

Visual Design: The Effect of Mere-Exposure in Different UX Phases

Marlene Vogel¹ and Johanna Silvennoinen²

Keywords: long-term user experience, UX phases, human-computer-interaction, visual user interface elements, design, mere-exposure

Abstract

To increase the understanding of long-term user experience (UX), we investigated the effect of mere-exposure and its influence on the evaluation of UX in different phases, e.g. pre-use, use, post-use and retrospectively after one week. A mere-exposure effect on the evaluation of beauty in a pre-use situation could be already detected in a former experiment by Vogel. In the present experiment two different graphic designs of a mobile application's user interface were tested. Visual attractiveness was manipulated to investigate a two-folded exposure effect. Pragmatic, hedonic product qualities and attractiveness as well as an overall liking have been assessed repeatedly. Concluding we could identify dynamics of UX. Additionally, UX phases differ also in influencing factors based on the varying importance of UX components. However, the mere-exposure of interfaces in a pre-use situation does not have any influence on the evaluation and UX of the same interfaces in the following UX phases.

Introduction

User experience (UX) [Alben, 1996; Hassenzahl, 2008] of interactive systems and services has been increasingly focused in product development and marketing [Desmet, 2002; Hassenzahl, 2005]. More recent models, assuming a dynamic character of the UX phenomenon [Minge, 2008; Kujala, Roto, Väänänen-Vainio-Mattila, Karapanos, & Sinnelä, 2011], describe different phases of UX [Karapanos, Zimmerman, Forlizzi, & Martens, 2009] and take the whole UX lifecycle into account [Pohlmeier, 2011]. This lifecycle includes anticipated as well as real, repeated, and memorized interaction with a product, system, or service. In order to facilitate positive long-term UX, these dynamics and related effects need to be considered when designing and evaluating UX.

To increase the understanding of the dynamic phenomenon of long-term UX [Kujala, Vogel, Pohlmeier, & Obrist, 2013] all phases need to be investigated. Pohlmeier [2011] describes several phases such as pre-use, use, post-use, repetitive use, past use with associated experiences, influencing factors and immediate as well as long-term effects. Anyway these assumptions have not been systematically

¹ Technische Universität Berlin, Research Group prometei, Germany

² University of Jyväskylä, Department of Computer Science and Information Systems, Finland

investigated yet. We strive to include the dynamics of repeated interactions and exposure in that framework of long-term UX. An effect that has consistently been found in this regard is the *mere exposure effect* [Zajonc, 1968, Bornstein, 1989], which we further examine in this paper. The *mere-exposure effect* known from social psychology could also be applied to the field of human-computer interaction (HCI) and therefore, needs to be investigated to gain more insights into the dynamics of UX. The mere-exposure effect predicts an increase of the liking or goodness-of-meaning of a target stimulus just by being exposed over and over to it. Additionally, Brickman, Redfield, Harrison, and Crandall [1972] identified a more specific and two-folded exposure effect depending on the first impression of a stimulus. Negatively evaluated stimuli were disliked more and more, whereas initially better liked stimuli were increasingly liked over the rate of exposure.

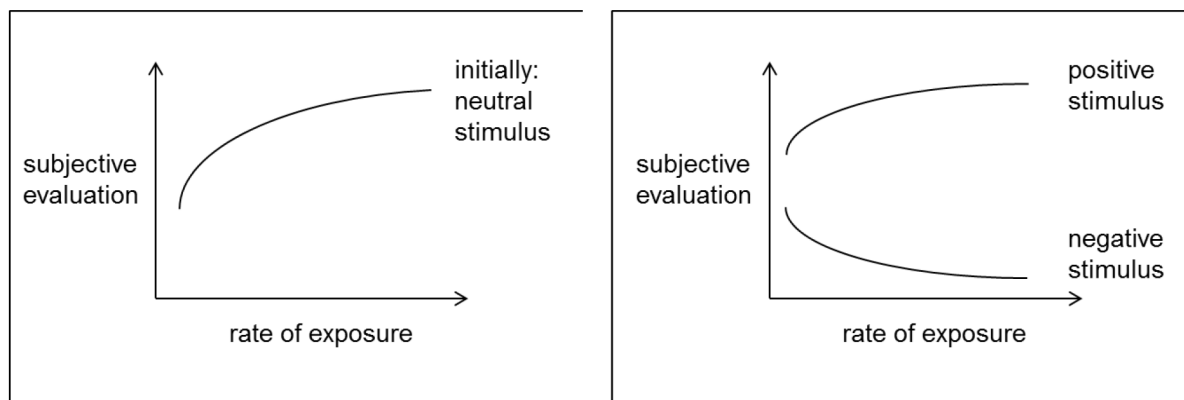


Fig. 1: Effects of mere-exposure. Left: Effect for first neutral evaluated stimulus. Right: Effect for first better and worse evaluated stimuli.

First results of this detailed and two-folded mere-exposure effect (Fig.1, right) in HCI could be detected by Vogel [2013] investigating the repeated evaluation of beauty of an interface in a pre-use situation. Based on these results influences of this effect on the evaluation of UX components [Thüring, & Mahlke, 2007] and experience in upcoming UX phases should be investigated.

Research Objective

This study investigates the mere-exposure of different interface designs (positive and negative version) in pre-use phases and possible effects regarding the UX in the following use and past-use phases. Focusing on the two-folded exposure effect caused by hedonic product qualities, two interface designs with high and low 'visual attractiveness' had been developed and evaluated before the experiment. It is known, that judgements and evaluations are more easily carried out when possibilities of comparing alternatives are provided [Saaty, & Vargas, 2012]. Because of this, the independent variable "visual attractiveness" was manipulated using a within-subject design, whereas "exposure rate" (control-group: low: $f=2$, treatment-group: high: $f=26$) was applied as a between-subject variable (Fig.3). Therefore, subjects were assigned randomly to one of the two groups: control-group

($N=17$) or treatment-group ($N=17$). Dependent variables were “liking” as well as “pragmatic” and “hedonic product qualities”, and “attractiveness”. They were assessed repeatedly (Fig. 3). 'Liking' was measured using a seven-point likert-scaled single-item. The abridged version of the AttrakDiff questionnaire ['AttrakDiff-mini' by Hassenzahl, & Monk, 2010] was used for measuring UX components. It includes ten seven-point semantic differential items, four to measure pragmatic quality, four to measure hedonic quality and two as measures of general product appreciation (attractiveness).

Stimulus Material

The stimulus material consisted of two interface design variations of a mobile application for transport search. The 'visual attractiveness' of these interfaces had been evaluated in a pre-test [Silvennoinen, Vogel, & Kujala, in preparation]. The most and least attractive versions out of four versions tested in the pre-test are serving as target-stimuli in the present experiment (Fig.2). 'Visual attractiveness' has been manipulated by two visual elements: colour (monochromatic vs. coloured) and perceived dimensionality of main functional elements (2D vs. 3D). Based on these elements the coloured 2D version (Fig.2, right) was seen as more visually attractive than the monochromatic 3D interface design (Fig.2, left).

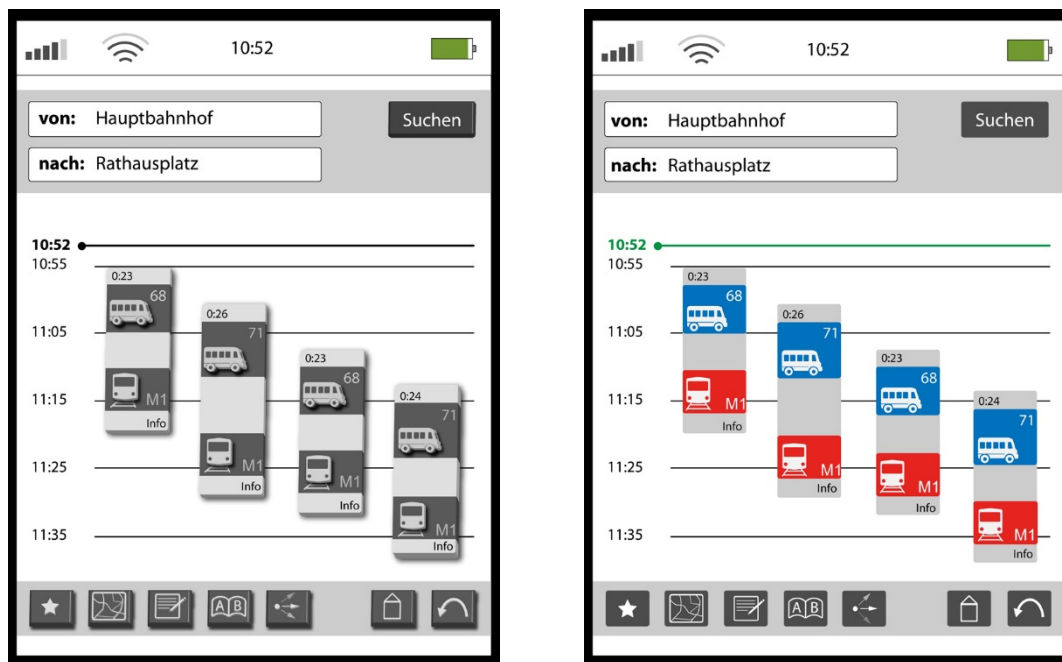


Fig. 2: Interfaces of a mobile application of transport search serving as target stimuli. Left: Low visual attractiveness design. Right: High visual attractiveness design.

The stimuli were presented on a 10" touch-sensitive display with a resolution of 768×1024 pixels. In the pre-use phase still pictures and in the use phase interactive simulations of these interface variations had been used.

Procedure & Participants

The experiment consisted of two parts. First, a laboratory session including the testing of pre-use and use phase. Second, an online questionnaire after one week to be answered from home by the participants testing the past use phase of the UX lifecycle (Fig. 3). Within the pre-use phase people first had to evaluate the interface designs (Fig. 2) on liking and subscales of the AttrakDiff-mini questionnaire ('pre I'). Subsequently target-stimuli were presented zero (control-group) or 24 times (treatment-group) in the 'exposure phase'. The targets were presented in a randomized order in six blocks with heterogeneous fillers. People had to concentrate on a secondary task and no further explicit evaluation of target stimuli was forced. Afterwards 'liking' and responses to the AttrakDiff-mini were recorded a second time ('pre II').

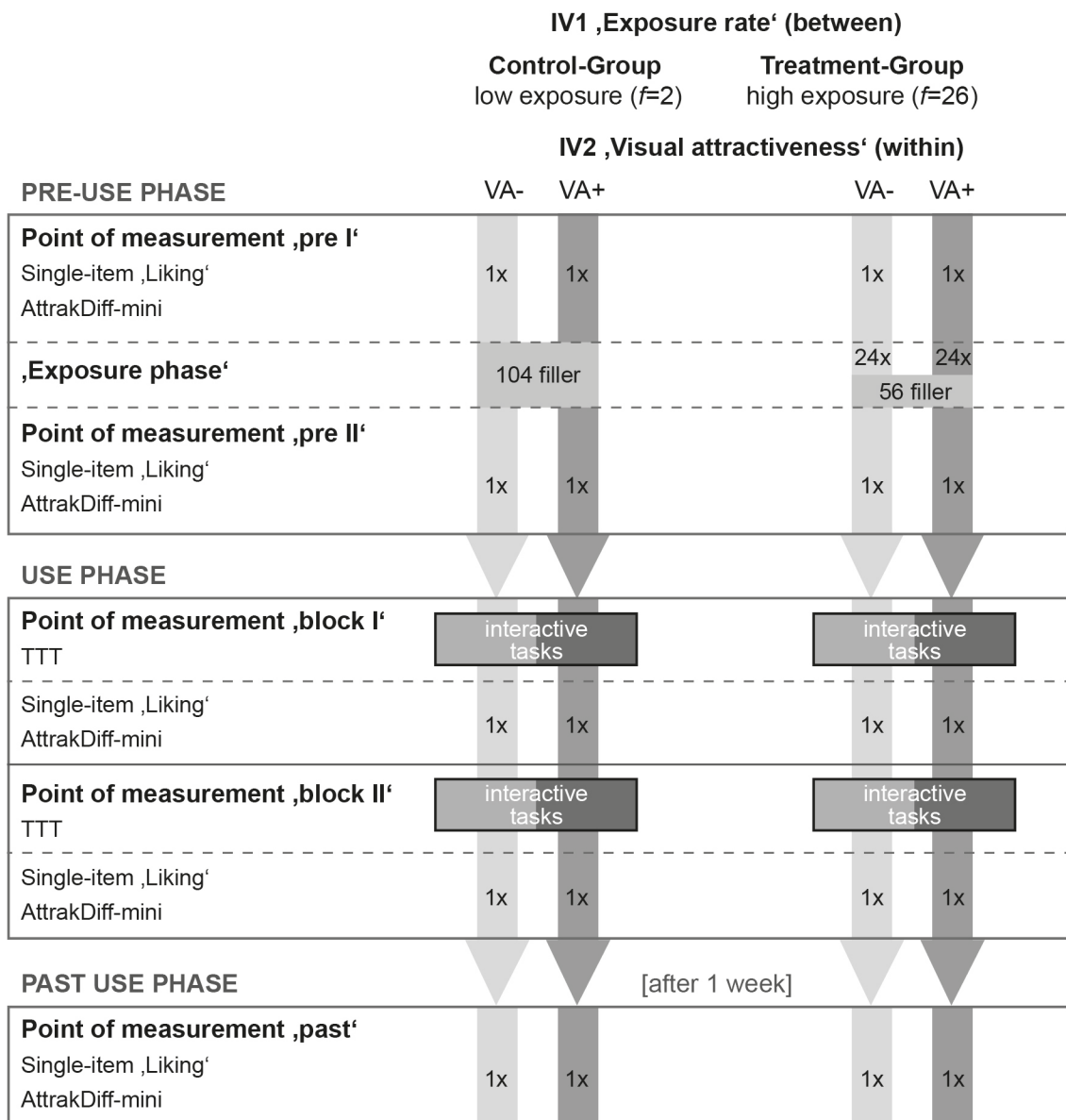


Fig. 3: Procedure and experimental design. IV = Independent Variable.

Within the use phase people had to interact and perform two different tasks in each block using each interface. Total task time (TTT in ms) was measured for each task. After each block of tasks ('block I' and 'II') people had to evaluate the interfaces again on 'liking' and AttrakDiff-mini scales. After one week each participant received a link for an online questionnaire where they had to rate the interfaces once more on 'liking' and AttrakDiff-mini ('past'). Demographic data as well as prior experience with similar applications were assessed for controlling purpose. Additionally the influence of preferences for aesthetical products has been measured by using the CVPA (Centrality of Visual Product Aesthetic) questionnaire [Bloch, Brunel, & Arnold, 2003].

34 people (15 female, 19 male, age: $M=28.1$, $SD=8.11$) participated voluntarily and received a reward of 15 Euro after completing the whole experiment including the online questionnaire after one week. Both treatment- and control-group consisted of 17 people.

Results

In this paper we concentrate on the analysis and discussion of the subjective evaluation and experience (Liking, UX components). Analysing the subjective evaluation the CVPA questionnaire has been included as a control variable. No main effects of 'visual attractiveness' (high/low) or 'exposure' (high/low) of the interfaces occurred in the pre-use phase. Regarding the mere-exposure effect a marginal significant interaction of 'visual attractiveness' (high/ low), 'exposure' (high/ low) and 'point of measurement' (pre I/ pre II) could be detected influencing 'general appreciation' (attractiveness) [$F(1,29)=3.095$, $p=0.089$, $part.\eta^2=0.096$] in the pre-use phase. People who received the high exposure showed the predicted two-folded mere-exposure effect, whereas people who did not receive extra exposure of the target stimuli showed the opposite pattern (Fig.4).

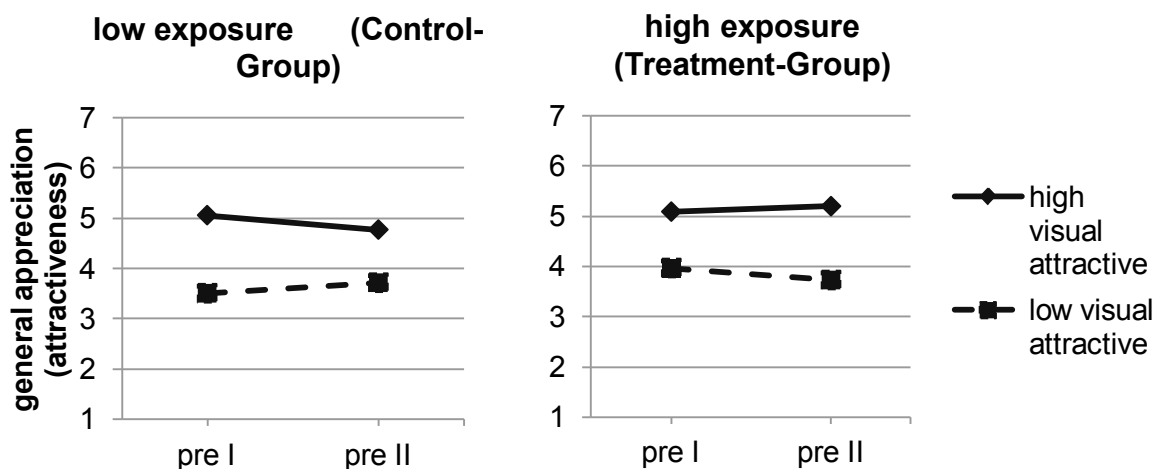


Fig. 4: Marginal significant interaction and influence of 'exposure' (low/high), 'visual attractiveness' (low/high) and 'point of measurement' (pre I and II) on 'general appreciation' (subscale 'attractiveness' of AttrakDiff-mini).

The analysis of the use phase show a significant interaction of 'exposure' (high/low) and 'point of measurement' (block I/ block II) and an influence on 'liking' [$F(1,31)=16.904$, $p<0.001$, $part.\eta^2=0.360$] (Fig.5) but without any differences of the two interface designs.

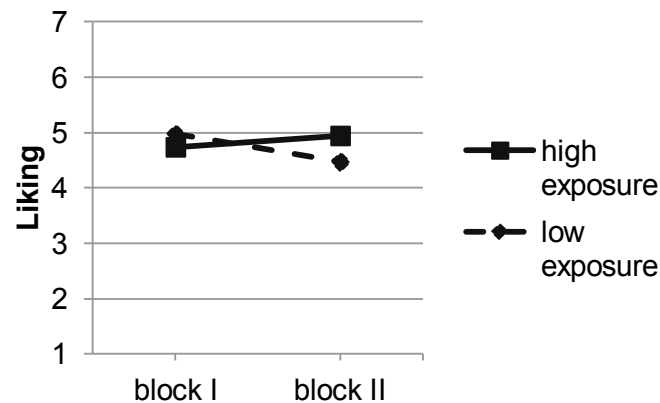


Fig. 5: Interaction of 'exposure' and 'point of measurement' and influence on 'liking' in the use phase.

Comparing and analysing the evaluation of the target stimuli right after the last interaction and after one week (past use phase) we could find main effects of the 'CVPA' (preference for visual product aesthetics) on 'hedonic product quality' [$F(1,30)=6.625$, $p=0.015$, $part.\eta^2=0.181$] and 'attractiveness' [$F(1,30)=10.104$, $p=0.003$, $part.\eta^2=0.252$]. Within a post-hoc analysis of the dichotomic CVPA factor, we could see that people with no preference (CVPA: $M<2.5$) for visual aesthetical products evaluated both interface variations higher on 'hedonic product quality' ($M=4.81$, $SE=0.26$) and 'attractiveness' ($M=5.78$, $SE=0.30$) than people with a preference (CVPA: $M\geq 2.5$) for visual aesthetical products ('hedonic product quality': $M=3.88$, $SE=0.09$, 'attractiveness': $M=4.64$, $SE=0.10$).

Discussion

To conclude we could find a marginal effect of exposure in the pre-use phase regarding the evaluation of attractiveness of two more or less attractive interfaces. Compared with the experimental design of Vogel (2013) the extra evaluation in the exposure phase seems to amplify the exposure effect. However, overall this effect does not influence the evaluation in the following phases of this kind of mobile application. Furthermore, we could show that the preference for visual product aesthetics (CVPA) has an influence on the evaluation of attractiveness and hedonic product qualities of this kind of mobile applications after an interaction and in the past use phase. For Practical implications the first impression and increasing of exposure of a product is significant for positive UX. This raises challenges and highlights the importance of well designed and aesthetically pleasing user interfaces already before an interaction takes place. In addition, if in the user interface

design process user preferences for visual product aesthetics has been taken into account more positive long-term UX is more likely to occur.

References

- Alben, L., (1996). Quality of experience: defining the criteria for effective interaction design. *Interactions*, 3, 11 – 15.
- Bloch, P. H., Brunel, F. F., & Arnold, T. J. (2003). Individual differences in the centrality of visual product aesthetics: concept and measurement. *Journal of consumer research*, 29(4), 551-565.
- Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, 106(2), 265–289.
- Brickman, P., Redfield, J., Harrison, A.A., & Crandall, R. (1972). Drive and predisposition as factors in the attitudinal effects of mere exposure. *Journal of Experimental Social Psychology*, 8(1), 31-44.
- Desmet, P. (2002). *Designing Emotions*. Doctoral Thesis, Delft University of Technology.
- Hassenzahl, M. (2005). Interaktive Produkte wahrnehmen, erleben, bewerten und gestalten. In: M. Eibl, H. Reiterer, P. F. Stephan, & F. Thissen (Hrsg.), *Knowledge Media Design – Grundlagen und Perspektiven einer neuen Gestaltungsdisziplin*, 151- 171, München: Oldenbourg.
- Hassenzahl, M., & Monk, A. (2010). The inference of perceived usability from beauty. *Human–Computer Interaction*, 25(3), 235-260.
- Karapanos, E., Zimmerman, J., Forlizzi, J., & Martens, J.B. (2009). User experience over time: an initial framework. In *Proc. CHI 2009*, 729–738, ACM Press.
- Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011). UX Curve: A method for evaluating long-term user experience. *Interacting with Computers*, 23, 473-483.
- Kujala, S., Vogel, M., Pohlmeier, A.E., & Obrist, M. (2013). Lost in Time: The Meaning of Temporal Aspects in User Experience. In *Proc. of CHI2013*. Paris, France.
- Minge, M. (2008). Dynamics of User Experience. Position paper for the workshop „Research Goals and Strategies for Studying User Experience and Emotion” at NordiCHI 2008, Lund, Sweden.
- Pohlmeier, A.E. (2011). *Identifying Attribute Importance in Early Product Development*. Technische Universität Berlin, Ph.D. thesis.
- Saaty, T. L. & Vargas L. G. (2012). How to Make a Decision. In Models, Methods, Concepts & Applications of the Analytic Hierarchy Process. *International Series in Operations Research & Management Science*, 175, 1-21.

- Thüring, M., & Mahlke, S. (2007). Usability, aesthetics, and emotions in human-technology interaction. *International Journal of Psychology*, 42 (4), 253-264.
- Vogel, M. (2013). Temporal Evaluation of Aesthetics of User Interfaces as one Component of User Experience. In Proc. of the Fourteenth Australasian User Interface Conference (AUIC2013), Adelaide, Australia. *Conferences in Research and Practice in Information Technology Series*, 139, 131-132, Australian Computer Society Inc.
- Zajonc, R.B. (1968): Attitudinal Effects of Mere Exposure. *Journal of Personality and Social Psychology*, 9(2), 1-27.

Einsatz von Schlüsselindikatoren für die Berücksichtigung unsicherer Randbedingungen in der Anlagenführung

Sebastian Werk¹

Schlüsselwörter: Schlüsselindikatoren, Unsicherheiten, Anlagenführung, Prozessindustrie, Wahrscheinlichkeitsnebenbedingungen

Zusammenfassung

Die Vorhersage dynamischen Prozessverhaltens ist auch für erfahrene Operateure oft nicht transparent. Gleichzeitig sind in der chemischen Industrie oft mehrere Parametern unsicher und fast alle Entscheidungen haben Einfluss auf die Profitabilität. Diese Faktoren machen es oft erforderlich, dass sich der Operateur zwischen einer riskanteren, profitableren und einer risikoärmeren, weniger profitablen Betriebsführung entscheiden muss.

Um dem Operateur in solchen Situationen eine bessere Entscheidungsgrundlage zu bieten, wurde ein Schlüsselindikator entwickelt, der Informationen über die Auswirkungen der Unsicherheiten enthält. Darauf aufbauend wurde die Hypothese, dass die Bereitstellung desselben zu besseren Entscheidungen führt, in einer ersten Studie in einer neu entwickelten Versuchsumgebung untersucht.

Innerhalb der Studie hat jene Gruppe, der der Schlüsselindikator zur Verfügung gestellt wurde, erst nach einer gewissen Lernphase signifikant bessere Entscheidungen getroffen als die Kontrollgruppe, der der Schlüsselindikator nicht zur Verfügung stand.

Einleitung

Die Überwachung und Steuerung chemischer Anlagen erfolgt in Leitwarten durch Operateure. Deren Aufgabe besteht darin, Prozesszustände und Regeleinrichtungen zu überwachen und bei Bedarf in die Betriebsführung einzugreifen.

Entscheidende Aspekte der Betriebsführung sind die in der Praxis oft existierenden *Unsicherheiten* verschiedenster Aspekte des Prozesses und zum anderen *konkurrierende Ziele*, deren Priorität der Operateur kontinuierlich abwägen muss.

Unsicherheiten betreffen sowohl die begrenzte Genauigkeit von Messstellen und Regelungsalgorithmen als auch das unvollständige Wissen um externe Einflussgrößen. So lassen sich oft Liefertermin und benötigte Menge eines Produktes nur begrenzt vorhersagen, auch unterliegen Ausgangsstoffe schwankender Qualität.

¹ Technische Universität Berlin, Graduiertenkolleg prometei, Fachbereich „Dynamik und Betrieb technischer Anlagen“. swerk@zmms.tu-berlin.de

Konkurrierende Ziele ergeben sich aus der simultanen Berücksichtigung von sicherheitsrelevanten und produktionsspezifischen Richtlinien auf der einen und Wirtschaftlichkeitsbetrachtungen auf der anderen Seite.

Im Wissen um Unsicherheiten und unter der Vorgabe konkurrierender Zielsetzungen stehen Operateure nun vor der Herausforderung, ihre Entscheidungen so zu treffen, dass sie einen Kompromiss aus einer sicheren, wirtschaftlich suboptimalen und einer riskanteren, wirtschaftlich profitableren Fahrweise darstellen.

Schlüsselindikator und Hypothesenentwicklung

Die Arbeit der Operateure wird dadurch erschwert, dass die Vorhersage dynamischen Prozessverhaltens auch für erfahrene Operateure nicht transparent ist (Su, 2004). In der Praxis liegen zunehmend dynamisch-algebraische Modelle vor, auf Basis derer die Anlage oder die Anlagenführung deterministisch optimiert wird. Die deterministische Auswertung führt bei Vorliegen von Unsicherheiten zwangsläufig zu einer begrenzten Aussagekraft der Ergebnisse. Aus diesem Grund wurden verschiedene stochastische Ansätze erforscht, die das Ziel haben, die Unsicherheiten im Rahmen der Optimierung zu berücksichtigen. (vgl. Kall & Wallace, 1994)

Ein solcher Ansatz ist die Verwendung von *Wahrscheinlichkeitsnebenbedingungen* (vgl. Wendt et. al., 2002; Gelatu et. al., 2010) entsprechend Ungleichung (1). Diese Nebenbedingungen stellen sicher, dass die definierten Zustände y^{\max} dem Prozessmodell f zufolge unter den äußeren Bedingungen u auch bei Vorliegen der Unsicherheiten ξ zumindest mit der Wahrscheinlichkeit α erreicht werden.

$$(1) \quad \Pr\{f(u, \xi) \geq y^{\max}\} \leq \alpha$$

In Werk et. al. (2012) wurde bereits die Idee vorgestellt, mit der Aussage der linken Seite der Gleichung die Entscheidungsfindung des Operateurs zu erleichtern. Inhaltlich steht sie verkürzt für „Wahrscheinlichkeit, einen definierten Zustand zu erreichen“ und ist beispielsweise als Prozentangabe darstellbar. Diese Aussage kann einen für das Operateursverhalten relevanten *Schlüsselindikator* darstellen.

Die in der Studie untersuchte Hypothese lautete somit: „Die Bereitstellung des Schlüsselindikators ‚Wahrscheinlichkeit, einen definierten Zustand zu erreichen‘ führt zu besseren Entscheidungen des Operateurs.“

Stichprobe

Insgesamt nahmen an der Studie 28 (♀: 11, ♂: 17) Probanden in der Kontrollgruppe und 25 (♀: 8, ♂: 17) in der Experimentalgruppe teil. Eine Person aus der Kontrollgruppe und zwei aus der Experimentalgruppe wurden von der Auswertung ausgeschlossen, da deren Ergebnis um mehr als die doppelte Standardabweichung von dem Mittelwert der jeweiligen Gruppe abwich. Dies war den Rückmeldungen nach

primär auf Verständnisprobleme hinsichtlich der Aufgabe zurückzuführen. Die Teilnahme wurde leistungsabhängig mit acht bis dreizehn Euro vergütet.

Versuchsumgebung

Zur experimentellen Überprüfung der zuvor entwickelten Hypothese ist eine Versuchsumgebung erforderlich, welche einen realitätsnahen Produktionsprozess mit einem Zielkonflikt abbildet. Weiterhin soll die Entscheidung des Operators von mehreren unsicheren Einflussgrößen abhängen, ohne die Aufgabe zu komplex werden zu lassen, um die Probanden nicht zu überfordern. Diese Versuchsumgebung wurde mangels geeigneter vorhandener selbst entworfen und umgesetzt.¹

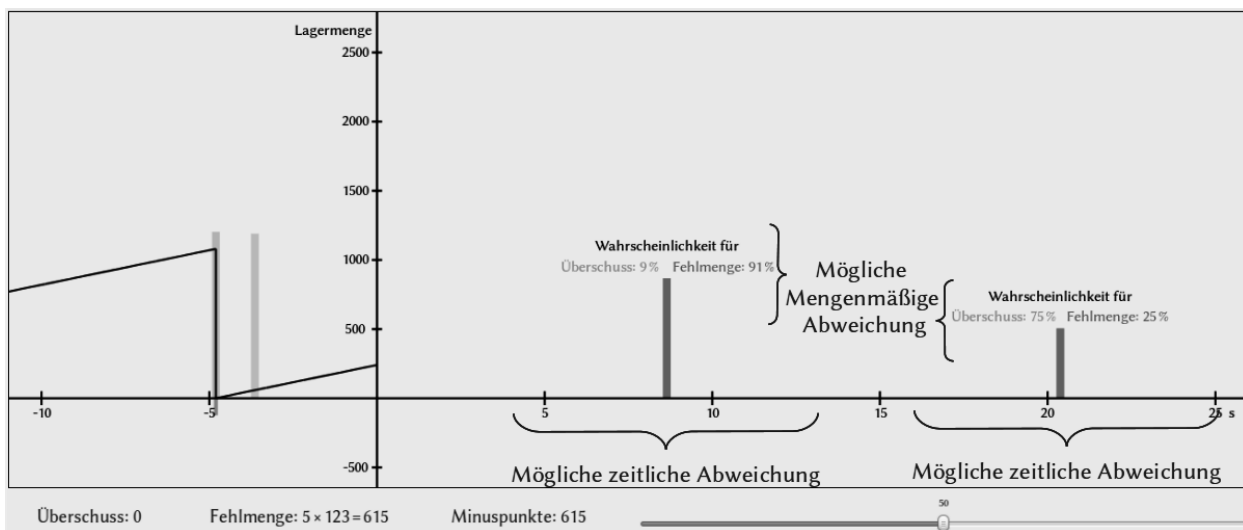


Abb. 1: Versuchsumgebung mit Schlüsselindikator. Geschweifte Klammern stellen mögliche zeitliche und mengenmäßige Abweichungen dar (für Probanden nicht direkt sichtbar). Prozentzahlen stehen nur Experimentalgruppe zur Verfügung.

Bei dem simulierten Prozess handelt es sich um folgende Dosieraufgabe: Alle 10 bis 15 Sekunden ist eine variierende Menge eines Stoffes erforderlich, welcher zu einem Vorprodukt hinzu gegeben wird. Da es sich bei diesem Stoff um einen Gefahrstoff handelt, darf er nicht produktionsnah gelagert werden, sondern muss zur Dosierung gepumpt werden. Der Operator muss nun die Pumpgeschwindigkeit so anpassen, dass die benötigte Menge zum vorgegebenen Zeitpunkt vorrätig ist.

Wird die benötigte Menge zum tatsächlichen Zielzeitpunkt übertroffen (Überschuss), so verletzt dies Sicherheitsbestimmungen und das Unternehmen muss Strafzahlungen leisten. Wird sie hingegen untertroffen (Fehlmenge), so ist das Produkt mangelhaft und muss aufwendig nachbearbeitet werden. Die Kosten je Kilogramm Fehlmenge sind fünf Mal so hoch wie jene je Kilogramm Überschuss.

¹ Die in Python umgesetzte Versuchsumgebung soll im kommenden Jahr als freie und offene Software veröffentlicht werden. Bei Interesse kann sie bereits zu Forschungszwecken zur Verfügung gestellt werden.

Der Zielkonflikt für die Probanden liegt somit darin, weder zu wenig noch zu viel zu produzieren. Entsprechend den verursachten Kosten erhält der Proband je Kilogramm Überschuss einen und je Kilogramm Fehlmenge fünf Strafpunkte.

Weiterhin wurde die Versuchsumgebung so umgesetzt, dass die Probanden zwar Mengen und Zeitpunkte sehen, dies jedoch nur Vorhersagen sind: Die tatsächlich benötigten Zeitpunkte und Mengen sind mit einer Standardabweichung von einer Sekunde beziehungsweise 15 Prozent um die Vorhersagen verteilt (siehe Abbildung 1). Die Probanden werden über diese Unsicherheiten informiert.

Bei manueller Änderung der Pumpenleistung stellt sich die gewünschte Pumpgeschwindigkeit erst mit Verzögerung ein. Dies ist ein bei der Steuerung von Prozessen häufig vorkommender Fall – bekannte Beispiele sind Kohlekraftwerke, die mehrere Stunden benötigen um aus dem Kaltstart Volllast zu erreichen. Die Anzeige für die Probanden wird stetig aktualisiert; in der Regel sehen sie die zuletzt erfüllte Dosieraufgabe sowie die zwei bis drei folgenden. (siehe Abbildung 1).

Die Experimentalgruppe verfügt zudem über den Schlüsselindikator. Dieser gibt für diesen Produktionsprozess an, mit welcher Wahrscheinlichkeit unter Berücksichtigung der Unsicherheiten bei jedem der gerade eingeblendeten Balken Überschuss oder Fehlmenge produziert werden.

Versuchsdesign

Es handelte sich um ein einfaktorielles Design. Der Faktor Gruppe war zwischen den Versuchspersonen zweifach gestuft in Experimental- und Kontrollgruppe. Die Experimentalgruppe verfügte im Gegensatz zur Kontrollgruppe über den Schlüsselindikator „Wahrscheinlichkeit, Überschuss (bzw. Fehlmenge) zu produzieren“.

Abhängige Variable

Zur Beurteilung der Leistung der Probanden wurden die erreichten Punktzahlen erfasst: Ein Punkt je Kilogramm Überschuss, fünf Punkte je Kilogramm Fehlmenge. Die Überprüfung auf statistisch signifikante Unterschiede zwischen den erreichten Punktzahlen der Gruppen fand im Rahmen eines t-Tests statt.

Durchführung

Die Probanden, maximal vier gleichzeitig, wurden begrüßt und zufällig einer Gruppe zugewiesen. Anschließend begannen sie, die Instruktionen der an einem normalen Bürorechner simulierten Leitwarte selbstständig in Form einer Bildschirmpräsentation durchzuarbeiten. Zunächst wurde hierbei der Schlüsselindikator nicht erwähnt, und Fragen waren zu keinem Zeitpunkt zugelassen.

Anschließend konnten sich die Probanden im Rahmen eines Übungsdurchgangs mit der Bedienung vertraut machen, im Zuge dessen sie zehn aufeinander folgende Dosieraufgaben bearbeiteten. Hierbei verfügte die Experimentalgruppe noch nicht

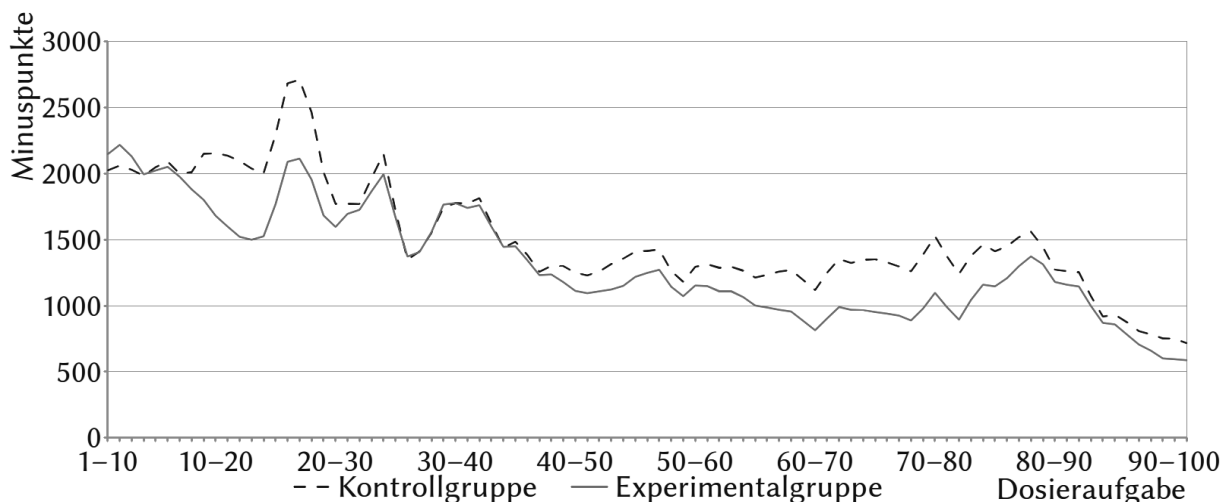
über den Schlüsselindikator. Die erreichten Punktzahlen wurden zur eventuellen Erkennung systematischer Unterschiede zwischen den Gruppen protokolliert.

Anschließend wurde den der Experimentalgruppe zugeordneten Probanden der Schlüsselindikator vorgestellt und inhaltlich erläutert. Nun erfolgte für alle Probanden der eigentliche Versuchsdurchgang mit einhundert aufeinander folgenden Dosieraufgaben; Probanden der Kontrollgruppe mussten sie ohne, jene der Experimentalgruppe mit dem Schlüsselindikator bearbeiten. Allen Probanden wurden die gleichen Aufgaben in gleicher Reihenfolge und gleichem Abstand gestellt.

Abschließend wurde die Arbeitsbelastung mit dem Fragebogen NASA-TLX (Hart & Staveland, 1988) erfasst und den Probanden Fragebögen zu Strategie, Verständnis und demografischen Daten vorgelegt.

Ergebnisse

Die durchschnittlichen Strafpunkte je Aufgabe der Experimentalgruppe ($M = 655$, $SD = 188$) lag unter der der Versuchsgruppe ($M = 744$, $SD = 233$), allerdings handelte es sich dabei nicht um einen signifikanten Unterschied, $t(48) = 1,45$,



$p = 0,15$.

Abb. 2: Gleitender Durchschnitt von jeweils zehn aufeinander folgenden Dosieraufgaben.

Die Punktzahlen unterschieden sich je nach Aufgabe relativ stark. Um den zeitlichen Verlauf besser einordnen zu können, ist in Abbildung 2 der gleitende Durchschnitt über jeweils zehn aufeinanderfolgende Dosieraufgaben dargestellt. Auffällig hierbei ist der bis etwa zur Mitte recht stark fallende und dann abflachende Verlauf. Möglicherweise handelt es sich hierbei um einen Lerneffekt.

Sollte dieser tatsächlich vorliegen, könnte ein Vergleich der durchschnittlichen Punkte nach Abschluss der Lernkurve aufschlussreich sein. Für eine erste Einschätzung wurde das Ende der Lernkurve dort vermutet, wo sich der gleitende Durchschnitt über fünf Aufgaben hin nicht mehr als ein Mal um mehr als zehn Prozent vom Vorgänger unterschied. Dies war mit der 53. Dosieraufgabe der Fall.

Für die Annahme, dass tatsächlich eine solche Lernkurve vorliegt, hat die Experimentalgruppe signifikant weniger Strafpunkte ($M = 484$, $SD = 159$) verursacht als die Kontrollgruppe ($M = 594$, $SD = 195$), $t(48) = 2,17$, $p = 0,04$.

Die Auswertung des Fragebogens zur Arbeitsbelastung hat keine signifikanten Unterschiede zwischen den beiden Gruppen ergeben. Der Schlüsselindikator wurde von den Probanden der Experimentalgruppe fast durchweg als nützlich erachtet: 13 Probanden beantworteten die Frage, ob sie ihn nützlich fanden, mit „Ja, sehr“, 10 mit „Ja, ein wenig“, keiner mit „Nein“ und einer mit „Nein, eher störend“. In den Fragebögen fanden sich schließlich noch eine Reihe von Hinweisen auf Schwächen und Fehler der Versuchsumgebung. Auf deren Inhalt und Folgen für Folgestudien wird im folgenden Abschnitt detailliert eingegangen.

Diskussion und Ausblick

Grundsätzlich kann festgehalten werden, dass in dieser Studie kein signifikant besseres Verhalten bei Bereitstellung des Schlüsselindikators festgestellt werden konnte. Die Entdeckung eines signifikanten Unterschieds nach einer gewissen Lernphase erfolgte rein explorativ ohne a priori Hypothese. Die Vermutung, dass Operateure nach einer gewissen Lernphase von dem Schlüsselindikator profitieren, sollte deshalb in einer Folgestudie genauer untersucht werden.

Diese Folgestudie befindet sich momentan in Vorbereitung. Sie wird neben der Kennenlernphase eine längere Übungsphase beinhalten. Weiterhin konnten aus den Fragebögen weitere wichtige Erkenntnisse hinsichtlich einer Verbesserung des Studiendesigns gewonnen werden. So sollten die Instruktionen zwecks eines besseren Verständnisses angepasst werden, die konkrete Pumpenleistung zur Vermeidung von extrem guten Ergebnissen durch Mathematik-affine Probanden entfernt und die Verzögerung der Pumpenleistung sowie der sichtbare Bereich vergrößert werden, um den Prozess realitätsnäher zu gestalten.

Literatur

- Geletu, A.; Hoffmann, A.; Klöppel, M. & Li, P. (2011). Monotony analysis and sparse-grid integration for nonlinear chance constrained process optimization. *Engineering Optimization*. 43, 1019-1041.
- Hart, S. G. & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) *Human Mental Workload*. Amsterdam: North Holland Press.
- Kall, P., & Wallace, S. W. (1994). *Stochastic. Programming*. Chichester: John Wiley & Sons.
- Su, H. T. (2004). *Operation-Oriented Advanced Process Control*. IEEE International Symposium on Intelligent Control. 252–257.

- Wendt, M.; Li, P. & Wozny, G. (2002). Nonlinear Chance-Constrained Process Optimization under Uncertainty. *Industrial & Engineering Chemistry Research*. 41, 3621-3629.
- Werk, S., Barz, T., Wozny, G., & Arellano-Garcia, H. (2012, 6). *An approach to process monitoring under probabilistic constraints*. Proceedings of the 22nd European Symposium on Computer Aided Process Engineering.

Analyse des Nutzungsverhaltens von Chirurgen beim Einsatz von multifunktionalen minimalinvasiven Chirurgieinstrumenten mit Handschaltern und Pedalen

Matthias Zickerow¹ und Marc Kraft²

Schlüsselwörter: Minimalinvasive Chirurgie, Chirurgische Instrumente, Gebrauchstauglichkeit, Handschalter, Pedale

Zusammenfassung

Um erfolgreich minimalinvasiv operieren zu können sind Hochfrequenz(HF)- oder Ultraschall(US)-Instrumente notwendig. Für beide Instrumentengattungen existieren Fußpedale und Handschalter als Bedienschnittstelle, jedoch wurden Vor- und Nachteile dieser bisher nicht abschließend untersucht. Daher wurde in der vorliegenden Beobachtungsstudie das Nutzungsverhalten von Chirurgen bei sechs Operationen dokumentiert und bzgl. dieser Benutzerschnittstellen ausgewertet. Es zeigte sich, dass HF- und US-Instrumente in mehr als 50 Prozent der Operationszeit eingesetzt wurden und auch mehr als die Hälfte aller Handlungen mit diesen Instrumenten ausgeführt wurden. Bei den ca. 2000 Handlungen kam es zu 15 durch die Benutzerschnittstelle bedingte Zwischenfälle. Dies waren 1,5 Prozent der Handlungen mit Instrumenten mit Handschalter und 0,8 Prozent mit Pedalen.

Einleitung

Die minimalinvasive Chirurgie bietet den Patienten erhebliche Vorteile, verlangt vom Operateur jedoch Einschränkungen in Bezug auf Sicht, Haptik und Bewegungsfreiheit (Breedveld, Stassen, & Meijer, 1999). Daher muss eine adäquate Blutstillung und eine möglichst kraftfreie Gewebetrennung möglich sein um erfolgreich minimalinvasiv operieren zu können. Denn nur so ist eine sichere Identifikation relevanter Gewebestrukturen ohne die Gefahr einer unbeabsichtigten Verletzung benachbarter Strukturen möglich.

Mit der HF- als auch mit der US-Technik sind solche blutarmen und kraftfreien Schnitte möglich. Beiden Techniken ist dabei gemein, dass der Energiefluss über einen von zwei Schaltern aktiviert werden muss. Diese Schalter werden aktuell als Fußpedale oder als Handschalter angeboten. Während bei wiederverwendbaren Instrumenten die Fußpedale dominieren sind die Mehrzahl der Einmalinstrumente mit Handschaltern für den Zeigefinger oder für Zeigefinger und Daumen ausgestattet. Dies führt dazu, dass mit der Hand eine Vielzahl von Instrumentenfunktionen

¹ Technische Universität Berlin, Graduiertenkolleg prometei

² Technische Universität Berlin, Fachgebiet Medizintechnik

ausgeführt werden müssen. Kleinfinger bis Mittelfinger dienen zum Schließen und meist auch Öffnen der Branche (als Gegenlager dient die Handinnenfläche oder der Daumen), der Zeigefinger bedient die Drehfunktion und – je nach konkreter Griffgestaltung – bedienen Zeigefinger und Daumen die Handschalter.

Bei der Nutzung von Pedalen konnten Einschränkungen der Gebrauchstauglichkeit nachgewiesen werden. Saucken et al. (2012) beobachteten und protokollierten 70 Operationen und registrierten dabei 10 Zwischenfälle, welche direkt mit der Pedalnutzung zusammen hingen. In einer Befragung an 74 Chirurgen gaben 58 Prozent an die Fußpedale häufig zu vertauschen, zudem lassen fünf Prozent der Chirurgen die Pedale auf Zuruf von nichtärztlichen OP-Personal bedienen. Trotz dieser Probleme würden nur 58 Prozent der befragten Chirurgen Handschalter gegenüber Pedalen bei gleichen Kosten bevorzugen. (Saucken, Donner, & Kraft, 2012)

Dies lässt vermuten, dass auch bei Handschaltern Probleme mit der Gebrauchstauglichkeit existieren, eine Studie welche dies aufzeigt ist jedoch nicht bekannt. Bereits 1999 beschrieben Breedveld et. al. (1999), dass es durch die Kombination der zwei Funktionen *Branchen Schließen* und *Positionieren* bei scherenartigen Instrumenten zu Ermüdung kommt. Desweiteren wies Berguer und Hreljac (2004) anhand 726 ausgewerteter Fragebögen einen signifikanten Einfluss der Handschuhgröße auf den prozentualen Anteil der Operationszeit nach, in dem die Chirurgen die Nutzung der Instrumente als schwierig empfanden.

Sowohl Breedveld et al. als auch Berguer et al. haben keine HF- oder US-Instrumente untersucht. Es scheint jedoch plausibel, dass sich die von Ihnen beschriebenen Probleme bei HF- und US-Instrumenten – durch mehr mit den Fingern zu bedienende Funktionen – verstärken. Bisher ermöglicht es aber keine Studie, einen qualitativen oder quantitativen Zusammenhang zwischen der Anzahl der Nutzungen einer Bedienschnittstelle und der Anzahl von Zwischenfällen herzustellen. Eine vergleichende Bewertung der Gebrauchstauglichkeit von Fußpedalen und Handschaltern bei chirurgischen Instrumenten ist daher zurzeit nicht möglich.

Ziel der Untersuchung ist es in einem ersten Schritt zur Bewertung der Gebrauchstauglichkeit von Handschaltern und Fußpedalen die Häufigkeit von Zwischenfällen zu untersuchen, dazu wurde ein Tool zur Dokumentation der Instrumentennutzung erstellt und bei laparoskopischen Operationen eingesetzt.

Methode

Es wurde die Instrumentennutzung von operierendem und assistierendem Chirurgen bei sechs laparoskopischen Operationen beobachtet und dokumentiert. Dabei wurde der medizinische Kontext sowie die Handlungen von nichtärztlichen OP-Personal, sofern es nicht direkt Einfluss auf die ärztlichen Handlungen hatte, zugunsten einer hohen zeitlichen Präzision der erfassten Nutzungen ausgespart. Als Nutzung wurden z.B. der Instrumentenwechsel, das Fassen von Gewebe sowie die Nutzung der HF- oder US-Funktionen dokumentiert; wenn es zu einer offensicht-

lich nicht beabsichtigten Interaktion mit dem Fußschalter oder Handschalter kam, wurde dies als Zwischenfall gewertet.

Material

Zur Dokumentation wurde ein Tablet mit 11,6 Zoll Touchscreen und eine in LibreOffice Base erstellter Benutzeroberfläche verwendet. Diese bietet über Drop-down Menüs für jede Hand des operierenden und des assistierenden Chirurgen mögliche Instrumente an. Je nach Instrumentenauswahl stehen in weiteren Menüs Handlungen mit dem Instrument zur Auswahl. Über Buttons können standardisierte Ereignisse, wie z.B. Geräte- oder Lichteinstellungen aber auch Zwischenfälle eingegeben werden. Sämtliche Eingaben werden in einer Datenbank abgelegt und mit einem Zeitstempel versehen.

Ergebnisse

Die Schnitt-Naht bzw. Schnitt-Konvertierungszeiten der dokumentierten Operationen variieren sehr stark und betragen zwischen 36min und 6h. Während der 11h und 14min dokumentierter Operationszeit wurde 1h und 36min (14 Prozent) mit HF-Instrumenten mit Fußschalter und 4h und 34min (41 Prozent) mit Ultraschallinstrumenten mit Handschaltern gearbeitet. Von den 2001 dokumentierten Handlungen von operierendem und assistierendem Chirurg waren 804 (40 Prozent) durch die Bedienung eines Instrumentes mit Handschalter bedingt, dabei traten zwölf Zwischenfälle (1,5%) auf. Bei den durch Instrumente mit Fußschaltern bedingten 374 (19 Prozent) Handlungen traten 3 (0,8 Prozent) Zwischenfälle auf.

Diskussion

Aufgrund der geringen Anzahl von Operationen und damit auch der beobachteten Chirurgen lassen sich diese Ergebnisse nicht grundsätzlich verallgemeinern, so sind z. B. 10 der 12 detektierten Zwischenfällen bei Handschaltern bei einer einzigen Operation aufgetreten.

Von Saucken et. al. (2012) dokumentierten in Ihrer Beobachtungsstudie bei 10 von 70 Operationen (14 Prozent) einen durch Pedale bedingten Zwischenfall, dies sind pro Operation deutlich weniger Zwischenfälle als bei der vorliegenden (3 Zwischenfälle bei 6 Operationen). Gleichzeitig ist die Anzahl der Zwischenfälle deutlich geringer als es die Umfrage von Saucken et. al. (2012) vermuten ließ, dort gaben 58 Prozent der befragten Chirurgen regelmäßige Verwechslungen der Pedale an.

Der festgestellte Unterschied zwischen der Selbsteinschätzung der Chirurgen und den dokumentierten Zwischenfällen könnte auf die geringe Rücklaufquote von 6,7 Prozent (Donner & Kraft, 2010) der Umfrage oder auf die schwierige Interpretierbarkeit der chirurgischen Handlungen während der Beobachtung hinweisen. Die Wirkung der beiden mittels Schaltern zu aktivierenden Funktionen ist

ähnlich und es besteht in den meisten Fällen keine eindeutige Indikation für die Nutzung einer der beiden Funktionen. So kann bei einer reinen Beobachtung ein Verwechseln der Schaltern nur durch eine Kompensationshandlung des Chirurgen erkannt werden, evtl. erfolgt diese aber nicht bei jeder Verwechslung. Dadurch würde die Anzahl an tatsächlichen Zwischenfällen bei einer Beobachtungsstudie, wie Sie hier vorliegt, unterschätzt werden. Für einen Vergleich beider Techniken hat dies wahrscheinlich keinen Einfluss, da dieser Effekt vermutlich bei Pedalen und Handschaltern gleichermaßen auftritt.

Aus den bisherigen Ergebnissen lässt sich nicht ableiten, weshalb eine Mehrzahl der Chirurgen (Saucken et al., 2012) Handschalter einem Pedal vorziehen würden.

Insgesamt zeigt sich aber, dass multifunktionale Chirurgieinstrumente während der Operation sehr lange (55 Prozent der gesamten Operationszeit) und sehr intensiv (59 Prozent der Handlungen) eingesetzt werden. Daher wird diese Studie fortgeführt und erweitert um einen Beitrag zur Bewertung der Gebrauchstauglichkeit von HF- und US-Instrumenten zu leisten.

Ausblick

Um die Gebrauchstauglichkeit der beiden Bedienkonzepte unter realen Bedingungen weiter zu untersuchen wird die Beobachtungsstudie weitergeführt. Diese wird durch Kurzinterviews der beobachteten Chirurgen ergänzt, um die festgestellten Probleme bei der Interpretierbarkeit der chirurgischen Handlungen zu reduzieren.

Um die Benutzerschnittstellen unabhängig von Trainingseffekten und weiteren Instrumentenbedingten Einflussfaktoren zu untersuchen wird eine Laborstudie mit Laien durchgeführt. Bei dieser wird die reale OP-Situation in einer für einen Laien verständliche Aufgabe nachgebildet und zu definierten Zeitpunkten eine Bedienung mittels Handschalter bzw. Pedal eingefordert.

Literatur

- Berguer, R., & Hreljac, A. (2004). The relationship between hand size and difficulty using surgical instruments: A survey of 726 laparoscopic surgeons. *Surgical Endoscopy*, 18(3), 508–512. doi:10.1007/s00464-003-8824-3
- Breedveld, P., Stassen, H. G., & Meijer, D. W. J. J. (1999). Manipulation in Laparoscopic Surgery: Overview of Impeding Effects and Supporting Aids. *Journal of Laparoendoscopic & Advanced Surgical Techniques*, 9(6), 469–480. doi:10.1089/lap.1999.9.469
- Donner, S., & Kraft, M. (2010). Potentials for an Improved Design of Instruments for Minimally-Invasive Surgery: Analysis of a Survey Conducted among Surgeons Specialized on Laparoscopic Procedures. In VDI Wissensforum (Ed.), *USEWARE. Nutzergerechte Gestaltung technischer Systeme* (pp. 179–188).

Saucken, A.-M. von, Donner, S., & Kraft, M. (2012). Ergonomic Problems Originating in the Use of High-Frequency and Ultrasonic Medical Devices. *Biomed Tech*, (57 (Suppl. 1)), 951–954.

Die Wirkung gerichteter Warnungen von Fahrerassistenzsystemen auf die Blickzuwendungs- und Reaktionszeiten von Autofahrern

Gert Weller¹, Fanny Heyne, Tim Feige, Hannes Bretschneider, Hendrik Oeser, Bernhard Schlag

Schlüsselwörter: Fahrerassistenzsysteme; multimodale, gerichtete Warnung; Fahrermodellierung

Zusammenfassung

Warnende Fahrerassistenzsysteme stellen eine Zwischenstufe zwischen rein informierenden und automatisch eingreifenden Systemen dar. Sie haben den Vorteil, dass der Fahrer im Zweifelsfall selbst entscheiden kann ob und wie er eingreift.

Zur Untersuchung der Frage, wie sich die Richtung einer Warnung auf die Reaktionszeiten auswirkt, wurden Probanden in einem Laborexperiment 18 Videosequenzen einer simulierten Fahrszene vorgespielt. In einigen dieser Sequenzen erfolgte die Darbietung einer kritischen Situation mit unterschiedlicher Warnmodalität, auf die mit Betätigung des Bremspedals reagiert werden musste. Ausgewertet wurden Blickzuwendungszeiten und Reaktionszeiten.

Neben einem positiven Effekt der Warnung im Vergleich zu keiner Warnung auf die Blickzuwendungszeit, zeigte sich hypothesenkonform eine signifikante Verringerung der Blickzuwendungszeiten und der Reaktionszeiten mit der gerichteten Warnung.

Theoretischer Hintergrund

Fahrerassistenzsysteme werden als ein probates Mittel zur Erhöhung der Verkehrssicherheit und zur Verringerung der Anzahl der Verkehrstoten ("towards zero", Wissenschaftlicher Beirat beim BMVBS, 2010) betrachtet. Generell muss bei Fahrerassistenzsystemen (FAS) zwischen informierenden, warnenden und eingreifenden Systemen unterschieden werden, wobei die Übergänge zwischen einzelnen Stufen oft fließend sind (Winner, Hakuli, & Wolf, 2012).

Warnende Systeme informieren den Fahrer vor einer drohenden Gefahr, ohne direkt einzugreifen. Sie haben somit den Vorteil, dass die Entscheidung ob und wie reagiert werden soll beim Fahrer liegt. Dies ist einerseits juristisch von Vorteil und hilft andererseits, den Fahrer „in-the-loop“ (also als aktiven Regler im Regelkreis) zu halten.

Gerade die Vorteile der warnenden Systeme sind auch deren Nachteile: der Mensch bleibt das schwächste Glied im Regelkreis, auf ihn gehen circa 90% aller

¹ Alle Autoren: Technische Universität Dresden, Professur für Verkehrspsychologie

Unfälle zurück. Somit muss bei der Gestaltung warnender Systeme besonders stark auf mögliche negative Auswirkungen geachtet werden. Insbesondere sind hier mangelnde Akzeptanz durch geringe Reliabilität der Warnung, zusätzliche Ablenkung durch die Warnung, sowie mögliche negative Verhaltensadaptation bis hin zum Kompetenzverlust zu nennen (Dingus, Jahns, Horowitz, & Knipling, 1997; Weller & Schlag, 2004; Wickens, Lee, Liu, & Becker, 2004).

Um negative Auswirkungen zu vermeiden, sollte die Warnung möglichst effektiv gestaltet werden. Bereits bekannt ist, dass multimodale Warnungen – üblicherweise in der Kombination aus visueller und akustischer Warnung – besonders effektiv sind (Kobiela, 2011). Offen bleibt dagegen, wie sich die Richtung der Warnung auf die Reaktionszeiten auswirkt. Legt man gängige Fahrerbremsmodelle zugrunde (etwa Zomotor, 1991, zitiert nach Kobiela, 2011) ist anzunehmen, dass sich durch gerichtete Warnungen – also Warnungen, die aus der gleichen Richtung wie die Gefahr kommen – die Reaktionszeiten weiter verringern lassen.

Für diesen Beitrag wurde untersucht, wie verschiedene Warnmodalitäten gerichtet und ungerichtet wirken. Dazu wurden die folgenden Hypothesen formuliert.

Hypothesen

Hypothese 1: Warnungen haben generell einen Einfluss auf die Reaktionszeit.

Hypothese 2: Warnungen haben generell einen Einfluss auf die Blickzuwendungszeit.

Hypothese 3: Multimodale Warnungen sind einfachen Warnungen überlegen, sie führen zu kürzeren Reaktionszeiten.

Hypothese 4: Multimodale Warnungen sind einfachen Warnungen überlegen, sie führen zu kürzeren Blickzuwendungszeiten.

Hypothese 5: Gerichtete Warnungen sind ungerichteten Warnungen überlegen, sie führen zu kürzeren Reaktionszeiten.

Hypothese 6: Gerichtete Warnungen sind ungerichteten Warnungen überlegen, sie führen zu kürzeren Blickzuwendungszeiten.

Methoden

Zur Untersuchung obiger Hypothesen wurde ein zweifaktorieller Versuchsplan mit Kontrollgruppe gewählt. Die Faktoren waren je zweifach gestuft.

Die Zuordnung der Probanden erfolgte randomisiert zu einer der fünf Versuchsbedingungen. Bei der Untersuchungsstichprobe handelt es sich um eine Gelegenheitsstichprobe, bestehend aus Studenten der TU Dresden und Bekannten der Versuchsleiter. Alle Probanden erhielten eine Aufwandsentschädigung. Insgesamt nahmen 60 Probanden an der Untersuchung teil, darunter 34 weibliche (Alter 18-48, $M=22$; $SD=7$) und 26 männliche (Alter 18-37, $M=22$; $SD=4$). Alle Probanden verfügten über eine gültige Fahrerlaubnis.

		R	
R	Faktor A Gerichtetheit	Faktor B Warnmodalität	
		visuell	Visuell-akustisch
		gerichtet	$P_{1,1}, Q_{1,1}$
	ungerichtet	$P_{2,1}, Q_{2,1}$	$P_{1,2}, Q_{1,2}$
	Ohne Warnung	P_0, Q_0	

aV P...Reaktionszeit

aV Q...Blickzuwendungszeit

Abbildung 6: 2x2-faktorielles Untersuchungsdesign mit Kontrollgruppe.

Jeder der fünf Gruppen wurden jeweils 18 Videos vorgeführt. Jeweils sechs Videos enthielten einen interessierenden Gefahrenreiz (im Folgenden Target genannt), sechs Filme enthielten irrelevante Gefahrenreize, sechs Videos enthielten keinerlei kritische Situationen und entsprechend keine FAS-Warnungen. Die letzten Videos wurden erstellt, um Erwartungseffekte zu minimieren. Die Targets bestanden aus Fußgängern, die von links (dreimal; ein Beispiel ist in Abbildung 7 dargestellt) oder von rechts (ebenfalls drei Videos) die Straße überquerten.

Die Dauer eines Videos betrug maximal 20 Sekunden. Der Inhalt der Videos wurde zunächst mit einer Simulationssoftware (STISIM-Drive™) als Simulation programmiert und dann mit FRAPS™, einer Screen-Capture-Software, aufgenommen. Die Warnungen wurden mit Adobe After Effects™ in die Videos so eingefügt, dass sie mit dem Frame beginnen, auf welchem der jeweilige Gefahrenreiz zum ersten Mal sichtbar ist. Je nach Situation erfolgte dies bei den Target-Reizen mit einer TTC zwischen 2,87 und 4,87 Sekunden. Die Situationen befanden sich alle innerorts, die Geschwindigkeit des Ego-Fahrzeuges wurde auf 50km/h festgelegt.

Den Probanden wurden die Videos auf einem Monitor präsentiert. An den Rechner war ein handelsübliches Lenkrad mit Pedalerie angeschlossen. Unterhalb des Monitors befand sich eine Vorrichtung zur Erhebung der Blickbewegung (System RED der Firma SMI). Vor Beginn des Versuches erfolgte die Kalibrierung der Blicke, diese wurde bei Bedarf zwischen einzelnen Videos wiederholt. Die Probanden wurden anschließend instruiert, das Video durch Drücken des Gaspedals zu starten und dieses solange gedrückt zu halten, bis eine Bremsung erforderlich ist. Sobald eine Gefahr im Video erkannt und gebremst wurde, wurde das Video beendet.



Abbildung 7: Ausschnitt aus einem Szenario (gerichtete visuelle Warnung; Target von links).

Für alle Videos mit Target wurden die Reaktionszeiten und die Blickzuwendungszeiten ausgewertet. Die Reaktionszeit wurde als diejenige Zeit definiert, die zwischen Beginn der Warnung und Beginn der Betätigung des Bremspedals verging. Die Blickzuwendungszeit wurde als diejenige Zeit definiert, die zwischen erstem Erscheinen der Warnung und dem ersten Anblicken des Target verging. Die Auswertung der Daten erfolgte in beiden Fällen mit der Software BeGaze 3.0 der Firma SMI, die statistische Berechnung der Daten mit SPSS 20.0.

Ausgehend von den Hypothesen, erfolgte die Auswertung der Daten jeweils mit einem t-Test für unabhängige Stichproben. Für die Hypothesen 1 und 2 (generelle Wirksamkeit einer Warnung im Vergleich zu ohne Warnung) wurden alle Bedingungen mit Warnung gemittelt und mit Hilfe eines t-Tests gegen die Bedingung ohne Warnung auf Unterschiede getestet. Für die anderen Hypothesen erfolgte die Auswertung ebenfalls mittels t-Test, die abhängigen Variablen wurden innerhalb einer Bedingung über alle sechs Targets gemittelt.

Ergebnisse

Vor der eigentlichen statistischen Auswertung auf Unterschiede zwischen den einzelnen Bedingungen, sind die Ergebnisse in Abbildung 8 und in Abbildung 9 grafisch dargestellt.

Die statistische Auswertung der Unterschiede zwischen der Bedingung ohne Warnung und den gemittelten Bedingungen mit Warnung ergab sich hinsichtlich der Reaktionszeit (Hypothese 1) keine signifikanten Unterschiede ($p=.061$; ohne Warnung: $M=1,47$; $SD=.15$; mit Warnung $M=1,27$; $SD=.37$).

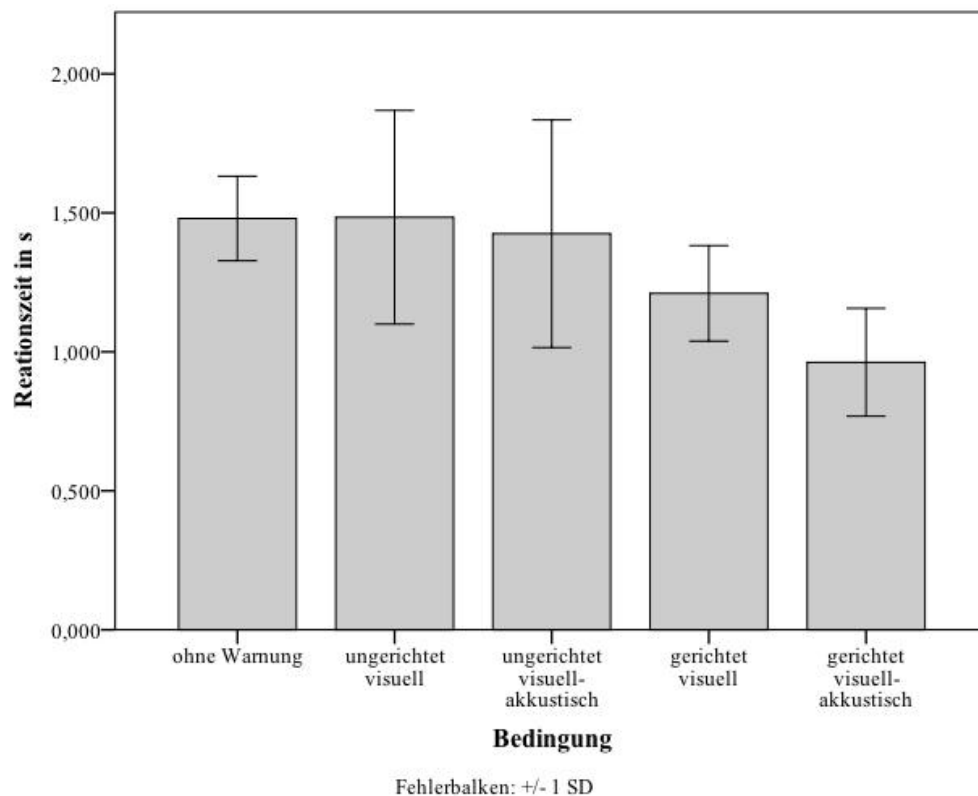


Abbildung 8: Veranschaulichung der mittleren Reaktionszeiten je Bedingung.

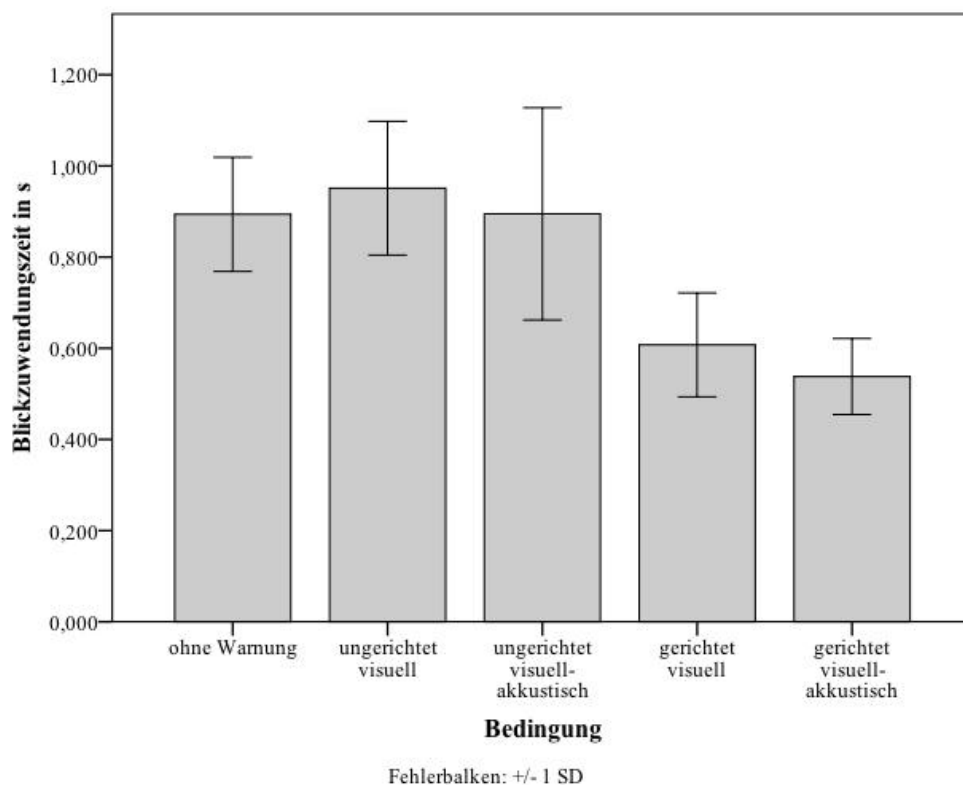


Abbildung 9: Veranschaulichung der mittleren Blickzuwendungszeit je Bedingung.

Die entsprechende Auswertung der Blickzuwendungszeiten (Hypothese 2) ergab dagegen einen hoch signifikanten Unterschied, wobei Blickzuwendungen mit

Warnungen schneller als ohne Warnungen erfolgten ($p < .05$; ohne Warnung: $M = 0,89$; $SD = .13$; mit Warnung $M = 0,75$; $SD = .24$).

Weiter wurde getestet ob sich multi-modale (visuell-akustisch) von einfachen (nur visuell) Warnungen unterschieden. Weder für die Reaktionszeiten (RT) (Hypothese 3) noch für die Blickzuwendungszeiten (BT) (Hypothese 4) ergaben sich signifikante Unterschiede (RT: $p = .09$; einfache Warnung: $M = 1,35$; $SD = .33$; multi-modale Warnung $M = 1,19$; $SD = .39$; BT: $p = .174$; einfache Warnung: $M = ,79$; $SD = .22$; multi-modale Warnung $M = ,72$; $SD = .25$).

Hinsichtlich der Hauptfragestellungen dieser Arbeit (Hypothesen 5 und 6), ergaben sich dagegen sowohl für die Reaktionszeiten als auch für die Blickzuwendungszeiten hoch signifikant niedrigere Werte mit gerichteter Warnung (RT: $p < .001$; ungerichtete Warnung: $M = 1,45$; $SD = .39$; gerichtete Warnung $M = 1,08$; $SD = .22$; BT: $p < .05$; ungerichtete Warnung: $M = ,92$; $SD = .19$; gerichtete Warnung $M = ,57$; $SD = .10$).

Diskussion

Es zeigte sich zunächst eine geringe Wirksamkeit der ungerichteten Warnungen (siehe Abbildung 8 und Abbildung 9). Im Gegensatz dazu war die gerichtete Warnung den Bedingungen ohne Warnung deutlich überlegen, unabhängig davon ob diese einfach oder multi-modal erfolgte.

Dieses Ergebnis zeigt deutlich das Potential warnender Assistenzsysteme. Werden diese gerichtet umgesetzt und gelingt es weiter, die Warnungen so reliabel zu gestalten, dass mögliche negative Effekte von FAS (siehe oben) ausgeschlossen werden können, haben warnende FAS ein deutliches Potential zur Verringerung der Unfallzahlen oder zumindest der Unfallschwere.

Trotz der deutlichen Ergebnisse soll abschließend noch auf methodische Schwächen des hier verfolgten Ansatzes eingegangen werden. Diese betreffen in erster Linie die Art der Erhebung und die Stichprobe.

Die Erhebung fand nicht im Realverkehr sondern im Labor statt. Dies erhöht zwar die interne Validität, gefährdet jedoch die externe Validität. Wünschenswert wäre hier die weitere Überprüfung der Ergebnisse im Simulator und folgend auch im Realverkehr. In diesem Zusammenhang sollte auch darauf geachtet werden, dass die Stichprobe ein annähernd reales Abbild der Fahrerpopulation darstellt. Denkbar wäre, dass ältere Fahrer spezifische Probleme mit Warnungen haben könnten.

Literatur

Dingus, T. A., Jahns, S. K., Horowitz, A. D., & Knipling, R. (1997). Human Factors Design Issues for Crash Avoidance Systems. In W. Barfield & T. A. Dingus (Hrsg.), *Human factors in intelligent transportation systems* (S. 55-93). Mahwah, NJ: Lawrence Erlbaum.

- Kobiela, F. (2011). Fahrerintentionserkennung für autonome Notbremssysteme. Wiesbaden: VS Verlag.
- Sivak, M. (1996). The information that drivers use: Is it indeed 90% visual? *Perception*, 25(9), 1081–1089.
- Weller, G., & Schlag, B. (2004). Verhaltensadaptation nach Einführung von Fahrerassistenzsystemen. In B. Schlag (Hrsg.), *Verkehrspsychologie. Mobilität – Verkehrssicherheit – Fahrerassistenz* (S. 351–370). Lengerich: Pabst.
- Wickens, C. D., Lee, J. D., Liu, Y., & Becker, S. E. G. (2004). *An Introduction to Human Factors Engineering* (2. Auflage). Upper Saddle River (NJ): Pearson Education.
- Winner, H., Hakuli, S., & Wolf, G. (Hrsg.). (2012). *Handbuch Fahrerassistenzsysteme. Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort* (2. korrigierte Auflage). Wiesbaden: Vieweg + Teubner.
- Wissenschaftlicher Beirat beim Bundesminister für Verkehr Bau und Stadtentwicklung. (2010). Sicherheit zuerst – Möglichkeiten zur Erhöhung der Straßenverkehrssicherheit in Deutschland. *Zeitschrift für Verkehrssicherheit*, 56(04), 171–194.

Auswirkungen warnender Fahrerassistenzsysteme auf das Blickverhalten und die Bremsreaktion abgelenkter Fahrer in Gefahrensituationen

Fanny Heyne¹, Daniela Deutsch¹, Gert Weller¹ und Bernhard Schlag¹

Schlüsselwörter: Fahrerablenkung, Fahrerassistenzsysteme, Reaktionszeit, Blickbewegung, Fahrermodellierung

Zusammenfassung

Ablenkung tritt im Straßenverkehr häufig auf und führt u.a. aufgrund der Beanspruchung begrenzter Aufmerksamkeitsressourcen zu einer Beeinträchtigung der Fahrleistung und einer Erhöhung des Unfallrisikos. Es wurde untersucht, wie sich die Ausführung verschieden komplexer visueller und kognitiver Nebenaufgaben auf das Blick- und Reaktionsverhalten von Fahrern in Gefahrensituationen auswirkt und inwiefern aufmerksamkeitslenkende, warnende Fahrerassistenzsysteme beides beeinflussen. Sechzig Probanden sahen an einem Computerbildschirm kurze Filmsequenzen und reagierten mittels Pedaldruck auf auftretende Gefahren. Dabei zeigte sich, dass Ablenkung generell zu einer verlangsamten Reaktion führt. Die Ausführung visueller oder besonders schwieriger Nebentätigkeiten verringert zudem die Anzahl korrekter Bremsungen. Die Wirksamkeit der FAS-Warnung wurde belegt, wobei hypothesenkonform gezeigt wurde, dass abgelenkte Fahrer stärker als nicht abgelenkte Fahrer profitieren: Ihr Nachteil, dass sie besonders viele Gefahren verpassen, wird durch die Warnung ausgeglichen. Die Ergebnisse fließen in ein Fahrermodell ein, welches gemeinsam mit der Professur für Kraftfahrzeugtechnik der TU Dresden entwickelt wurde.

Ausgangslage

Gefährliche Situationen im Straßenverkehr verlangen ein schnelles und richtiges Eingreifen des Fahrers. Neben der Möglichkeit des Ausweichens stellt das Bremsen eine der wichtigsten Handlungsalternativen in Gefahrensituationen dar. Die Bremsreaktionszeit setzt sich additiv aus mehreren einzelnen Zeitdauern zusammen, welche für sequentiell ablaufende Teilprozesse der Reaktion benötigt werden (z.B. Bäumler, 2002; Burckhardt, 1980, 1985; Green, 2000; Löhle, 1983). Neben der muskulären Reaktion sind zunächst verschiedene kognitive Verarbeitungsschritte (Wahrnehmen, Erkennen, Entscheiden) notwendig, für die wiederum von einigen Autoren eine Blickzuwendung zur Identifikation des Reizes als Voraussetzung angesehen wird (z.B. Bäumler, 2002; Burckhardt, 1985).

Es gibt verschiedene Ursachen für eine fehlende oder zu langsame Reaktion. Eine ist die zu hohe kognitive (Green, 2000) bzw. visuelle Belastung (Summala,

¹ Technische Universität Dresden, Professur für Verkehrspsychologie

2000), wie sie bspw. durch Ablenkung entsteht. Im Verkehr tritt Ablenkung auf, wenn der Fahrer seine Aufmerksamkeit nicht auf das Fahren, sondern auf eine zweite, konkurrierende Aufgabe innerhalb oder außerhalb des Fahrzeugs richtet, welche nicht direkt mit der Fahraufgabe in Verbindung steht (Regan, Hallett und Gordon, 2011). Huemer und Vollrath (2011, 2012) zeigen, dass Fahrzeugführer ca. 30-50% der Fahrtzeit mit der Ausführung fahrfremder Tätigkeiten verbringen. Diese beanspruchen die begrenzten Aufmerksamkeitsressourcen des Fahrers und beeinträchtigen die Fahrleistung so, dass die Fahraufgabe nicht mehr sicher ausgeführt werden kann. Negative Auswirkungen zeigen sich z.B. in der Lenkung, der Spurhaltung und der Fahrzeugkontrolle (u.a. Liang & Lee, 2000). Des Weiteren wirkt sich Ablenkung auf die Wahrnehmung von Gefahren aus und führt zu einer Verlängerung der Reaktionszeit (u.a. Alm & Nilsson, 1994; Lamble, Kauranen, Laakso & Summala, 1999; Strayer & Johnston, 2001). Entsprechend stellt Fahrerablenkung eine der wichtigsten Unfallursachen dar. Sie trägt zu circa 22% aller Unfälle bei und erhöht das individuelle Risiko zu verunfallen um das bis zu 5,5-Fache im Vergleich zu einer Fahrt ohne Ablenkung. Komplexe Nebentätigkeiten erhöhen das Unfallrisiko dabei stärker, als weniger komplexe Tätigkeiten (Klauer, Dingus, Neale, Sudweeks & Ramsey, 2006).

Zur Vermeidung von Unfällen wurden u.a. warnende Fahrerassistenzsysteme (FAS) entwickelt. Diese unterstützen den Fahrer in Gefahrensituationen, indem sie seinen Aufmerksamkeitsfokus auf die Gefahr lenken, so dass er kollisionsvermeidende Reaktionen ausführen kann (Lee, Ries, McGehee & Brown, 2000). Die Wirksamkeit solcher Systeme wurde in zahlreichen Studien belegt (z.B. Doisl, 2004; Fricke, 2009; Lee et al., 2000).

Vorliegende Arbeit untersucht Besonderheiten im Reaktions- und Blickverhalten abgelenkter Fahrer in Gefahrensituationen und prüft die Annahme, dass abgelenkte Fahrer stärker von warnenden FAS profitieren als nicht-abgelenkte Fahrer.

Methodik

An der Untersuchung nahmen 60, vorwiegend studentische Probanden (61,7% weiblich) teil. Diese waren im Mittel 24,3 Jahre alt (SD: 5,0) und besaßen seit 6,6 Jahren einen Führerschein (SD: 4,9).

Die Aufgabe der Probanden bestand darin kurze Filmsequenzen anzuschauen, die auf einem Computerbildschirm dargeboten wurden, und mittels Druck auf ein Bremspedal auf auftretende Gefahren zu reagieren. Parallel sollten sie visuelle bzw. kognitive Nebenaufgaben ausführen, wobei die Stärke der von diesen Aufgaben ausgehenden Ablenkung als within-subject Faktor in ausbalancierter Reihenfolge von „nicht vorhanden“ über „gering“ bis „stark“ variierte. Die Hälfte der Studienteilnehmer wurde durch ein akustisches Signal gewarnt, sobald in den Filmsequenzen eine Gefahrensituation entstand. Die andere Hälfte der Probanden erhielt keine Warnung. Insgesamt ergab sich daher ein 2x2x3-faktorielles Untersuchungsdesign mit Messwiederholung auf dem Faktor „Stärke der Ablenkung“. Als

abhängige Variablen wurden die Anzahl korrekter Bremspedalbetätigungen bei Auftauchen einer Gefahr, die Bremsreaktionszeit und verschiedene Blickmaße erhoben.

Das Untersuchungsmaterial (vgl. Abb.1) bestand aus 30, mit Hilfe des VICOM-Editors der TÜV | DEKRA arge tp 21 erstellten, kurzen (16s - 67s) Filmsequenzen, welche verschiedene Fahrsituationen jeweils aus Sicht des Pkw-Fahrers darstellten. Um der Erwartung der Probanden entgegenzuwirken, enthielten nur 12 der Filme die interessierende Gefahrensituation „querender Fußgänger“. Die kognitive Nebenaufgabe bestand aus Additions- (geringe Ablenkung) bzw. Subtraktionsaufgaben (starke Ablenkung), welche den Probanden vor jeder Filmsequenz genannt wurden und fortlaufend mit dem Ergebnis der letzten Rechnung weitergeführt werden sollten. Unter der Bedingung visueller Ablenkung wurde ein Suchbild im Bereich des rechten Rückspiegels eingeblendet. Die Aufgabe der Probanden bestand darin, ein braunes „X“ oder ein blaues „O“ zu finden. Es war jeweils genau einer der beiden Zielreize zu sehen, wobei sich dieser unter der Bedingung geringer Ablenkung durch ein Merkmal (Farbe oder Form; Merkmalssuche) und unter der Bedingung starker Ablenkung durch beide Merkmale (Farbe und Form; serielle Suche nach Treisman & Gelade, 1980) von den 18 Distraktoren unterschied. Unabhängig davon, ob die Probanden die Aufgabe richtig gelöst hatten, wurde aller vier Sekunden eine neue Suchaufgabe eingeblendet.



Abb. 1: Untersuchungsmaterial. Links: Ausschnitt aus einer Sequenz mit dem Gefahrenreiz „Fußgänger von rechts“. Rechts: Beispiele für die visuelle Nebenaufgabe (oben: geringe Ablenkung, unten: starke Ablenkung).

Erste Ergebnisse

Haupteffekt „FAS-Warnung“. Wurden die Probanden gewarnt reagierten sie im Mittel nach 721ms (SD=29) auf das Erscheinen eines Fußgängers. Wurden sie nicht gewarnt erfolgte die Reaktion nach 892ms (SD=31). Die Warnung führte somit zu einer signifikanten Reduktion der Reaktionszeit um ca. 170ms

($F(1)=16,049$, $p=.000$). Zudem wurden unter der Bedingung mit Warnung mehr korrekte Bremsungen ausgeführt als ohne Warnung ($F(1)=67,959$, $p=.000$). In den Blickmaßen konnte kein Effekt des Faktors „FAS-Warnung“ nachgewiesen werden.

Haupteffekt „Art der Ablenkung“. Wurden die Probanden mit einer visuellen Nebenaufgabe abgelenkt war die Anzahl der korrekten Bremsungen bei Erscheinen eines Fußgängers geringer, als bei der Ausführung der kognitiven Nebenaufgabe ($F(1)=7,299$, $p=.009$). Die Reaktionszeit unterschied sich nicht signifikant zwischen den Ablenkungsbedingungen. Die Blickmaße zeigen, dass der Fußgänger bei visueller Ablenkung später (aV „Zeit bis zur ersten Fixation“, $F(1)=8,311$, $p=.006$), seltener (aV „Anzahl Fixationen“, $F(1)=4,309$, $p=.044$) und kürzer (aV „prozentuale Verweildauer“, $F(1)=4,083$, $p=.05$) betrachtet wurde, als bei der Ausführung der kognitiven Nebenaufgabe.

Haupteffekt „Stärke der Ablenkung“. Die Stärke der Ablenkung wirkte sich sowohl auf die Anzahl korrekter Bremsungen ($F(1,650)=47,298$, $p=.000$) als auch auf die Reaktionszeit ($F(2)=66,817$, $p=.000$) aus. Waren die Probanden stark abgelenkt zeigten sie signifikant weniger korrekte Bremsungen als unter den Bedingungen geringer ($F(1)=77,125$, $p=.000$) und fehlender Ablenkung ($F(1)=47,668$, $p=.000$). Zudem reagierten stark abgelenkte Probanden mit einer Reaktionszeit von 890ms ($SD=21$) rund 280ms langsamer als nicht abgelenkte Versuchsteilnehmer ($F(1)=94,254$, $p=.000$). Eine geringe Ablenkung ($M=922$ ms, $SD=25$) führte zu ähnlichen Zeitverlusten im Vergleich zur Kontrollbedingung ohne Ablenkung, in der im Mittel bereits nach 606ms ($SD=35$) reagiert wurde ($F(1)=94,024$, $p=.000$). Der Effekt der Ablenkungsstärke zeigte sich zudem in der Zeit bis zur ersten Fixation ($F(2)=4,817$, $p=.010$), der Anzahl der Fixationen ($F(2)=5,654$, $p=.005$) und in der prozentualen Verweildauer ($F(2)=9,428$, $p=.000$). Waren die Probanden nicht abgelenkt fixierten sie den Fußgänger signifikant früher als bei geringer ($F(1)=9,935$, $p=.003$) bzw. starker Ablenkung ($F(1)=4,488$, $p=.040$). Der Fußgänger wurde unter der Bedingung ohne Ablenkung häufiger fixiert als unter den Bedingungen geringer ($F(1)=11,246$, $p=.002$) und starker Ablenkung ($F(1)=4,961$, $p=.031$) und die prozentuale Verweildauer war erhöht (keine vs. geringe Ablenkung: $F(1)=6,852$, $p=.012$; keine vs. starke Ablenkung: $F(1)=17,103$, $p=.000$).

Interaktionseffekte. Art und Stärke der Ablenkung wirkten sich in Abhängigkeit des Faktors „FAS-Warnung“ unterschiedlich auf die Anzahl der korrekten Bremsungen aus. In der Bedingung mit visueller Ablenkung erhöhte sich diese durch die FAS-Warnung stärker, als in der Bedingung mit kognitiver Ablenkung ($F(1)=15,924$, $p=.000$; vgl. Abb. 2). Die Interaktion zwischen den Faktoren „FAS-Warnung“ und „Stärke der Ablenkung“ ($F(1,650)=37,564$, $p=.000$) wird durch Abbildung 2 verdeutlicht. Unter der Bedingung mit starker Ablenkung erhöhte sich die Anzahl der richtigen Bremsungen durch die FAS-Warnung signifikant stärker als in den Bedingungen mit geringer ($F(1)=7,035$, $p=.011$) bzw. ohne Ablenkung ($F(1)=8,341$, $p=.006$).

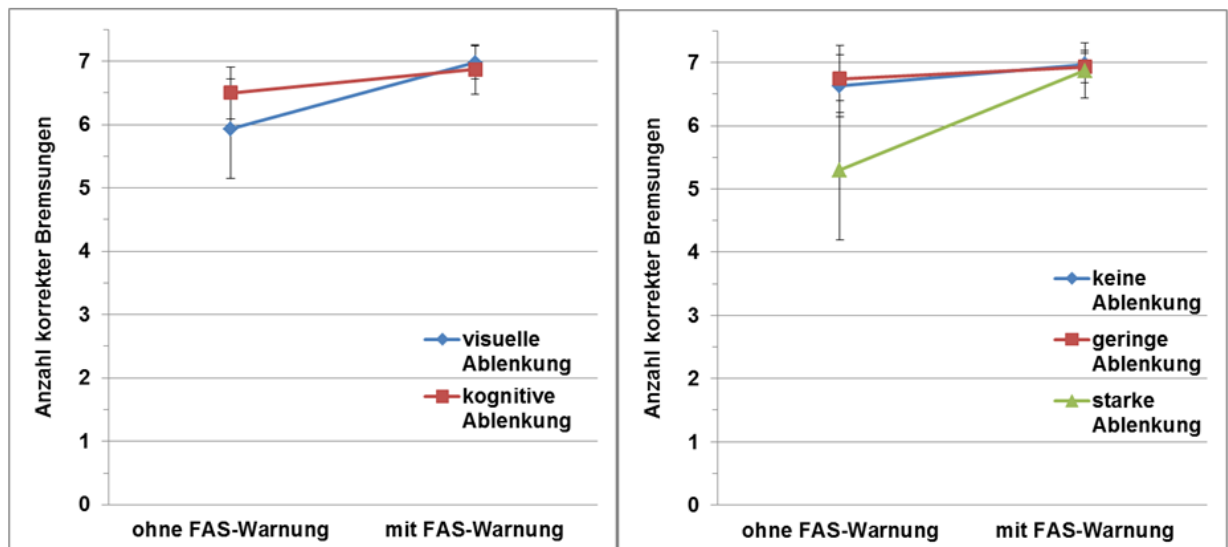


Abb. 2: Interaktionen zwischen der FAS-Warnung und der Art der Ablenkung (links) bzw. der Stärke der Ablenkung(rechts) für die aV „Anzahl korrekter Bremsungen“.

Diskussion und Ausblick

Die Zunahme korrekter Bremsungen und die Abnahme der Reaktionszeit durch die FAS-Warnung stehen in Einklang mit früheren Befunden und sprechen für die Wirksamkeit von warnenden Fahrerassistenzsystemen. Interessant ist jedoch, dass sich der positive Effekt der FAS-Warnung nur in den Reaktions-, nicht jedoch in den Blickmaßen zeigte. Ausgehend von der Annahme, dass eine Warnung die Aufmerksamkeit des Fahrers auf die Gefahr lenkt, wurde erwartet, dass die Probanden in der Bedingung mit FAS-Warnung den Fußgänger schneller entdecken. Der fehlende Effekt der FAS-Warnung auf die Zeit bis zur ersten Fixation deutet aber darauf hin, dass durch die Warnung nicht die Blickzuwendung zur Gefahr, sondern ein anderer Teilprozess der Reaktion (vgl. Burckhardt, 1985) beschleunigt wurde. Eine weitere mögliche Erklärung ist, dass die Probanden unter der Bedingung mit FAS-Warnung das Bremspedal bereits betätigten, ohne den Gefahrenreiz fixiert und identifiziert zu haben. Zöller und Hugemann (1998) gehen bspw. davon aus, dass eine Reaktion auch ohne Blickzuwendung stattfinden kann. Sie unterscheiden, in Anlehnung an Donders (1869/1969), zwischen der Detektion des Gefahrenreizes und der Identifikation. Während letztere die Voraussetzung für die Ausführung einer Wahlreaktion oder einer Go-NoGo-Reaktion ist, reicht für eine einfache Reaktion die reine Detektion des Gefahrenreizes aus, für die jedoch keine Blickzuwendung notwendig ist. Eine Blickzuwendung hin zu einem peripheren Gefahrenreiz ist somit für eine einfache Detektion, nicht notwendig. Eventuell führte die FAS-Warnung dazu, dass aus einer Wahlreaktion eine einfache Reaktion wurde und eine Blickzuwendung somit entfallen konnte. Weitere Auswertungen der Daten dieser Studie und einer bereits durchgeführten Studie im Fahrsimulator werden zur Ursache der verkürzten Reaktionszeit durch die FAS-Warnung zusätzliche Erkenntnisse liefern.

In Einklang mit früheren Befunden zeigte sich, dass Ablenkung die Reaktion auf Gefahrenreize verlangsamt. Dieser Effekt besteht unabhängig davon wie stark die Ablenkung ist oder ob es sich um visuelle oder kognitive Ablenkung handelt.

Visuelle Ablenkung beeinträchtigt die Reaktion auf Gefahren jedoch stärker als kognitive Ablenkung, was sich in einer signifikant geringeren Anzahl korrekter Bremsungen zeigte. Dieses Ergebnis steht in Einklang mit den Blickmaßen, die unter der visuellen Bedingung eine seltenere und spätere Entdeckung des Gefahrenreizes belegen. Eine Erklärung dafür ist, dass die visuelle im Gegensatz zu der kognitiven Nebenaufgabe eine Blickabwendung verlangt. Je weiter ein Gefahrenreiz vom aktuellen Fixationsort entfernt ist, desto geringer ist die Wahrscheinlichkeit, dass die visuelle Aufmerksamkeit auf ihn gerichtet wird (SEEV, Wickens & Horrey, 2008). Auch basierend auf der Theorie der multiplen Ressourcen (Wickens, 1984) ist davon auszugehen, dass die vorwiegend durch visuelle Informationsaufnahme geprägte Fahraufgabe besonders stark durch eine ebenfalls die visuelle Modalität beanspruchende Nebenaufgabe beeinträchtigt wird.

Für den Faktor „Stärke der Ablenkung“ stimmen die Ergebnisse der Blickmaße mit den Befunden zur Reaktionszeit überein. Sie zeigen, dass Ablenkung generell – unabhängig von ihrer Stärke – dazu führt, dass ein Gefahrenreiz später entdeckt und seltener und kürzer angeschaut wird. Die spätere Entdeckung erklärt nach dem Modell von Burckhardt (1985) die verlängerte Gesamtreaktionszeit. Die verkürzte Verweildauer auf den Gefahrenreiz kann als Hinweis für die erhöhte Beanspruchung durch die Nebenaufgabe interpretiert werden: Bei zeitkritischen Aufgaben, wie beispielsweise der Fahrzeugführung, äußert sich zunehmende Beanspruchung in einer Verringerung der Fixationsdauer (Joos, Rötting & Velichkovsky, 2003), und sollte somit ähnlich auf das verwandte Maß der Verweildauer wirken. Anhand der korrekten Bremsreaktionen zeigte sich zudem, dass starke Ablenkung die Reaktion auf Gefahren stärker beeinträchtigt, als geringe Ablenkung. Dieses Ergebnis steht in Einklang mit den Erkenntnissen zum Unfallrisiko (vgl. Klauer et al., 2006) und lässt sich ebenfalls anhand bekannter Ressourcentheorien erklären, die von einer Begrenzung der zur Verfügung stehenden Verarbeitungskapazität ausgehen.

Die Interaktionseffekte zeigen, dass sich die Anzahl korrekter Bremsreaktionen der visuell bzw. stark abgelenkten Probanden durch die FAS-Warnung jeweils an die Anzahl korrekter Bremsreaktionen der kognitiv bzw. nicht oder nur gering abgelenkten Probanden angleicht. Mit FAS-Warnung erreichen alle Gruppen ein hohes Niveau (vgl. Abb. 2; max. Anzahl korrekter Bremsreaktionen im Versuch = 7). Insgesamt profitieren abgelenkte Fahrer also besonders stark von warnenden FAS, da sie durch diese erstens – wie alle Fahrer – eine schnellere Reaktionszeit aufweisen und sich bei ihnen zweitens die Anzahl korrekter Reaktionen –ausgehend von einem niedrigen Ausgangswert – besonders stark erhöht.

Literatur

- Alm, H., & Nilsson, L. (1994). Changes in driver behaviour as a function of hands-free mobile phones - A simulator study. *Accident Analysis and Prevention*, 26(4), 441-451. doi: 10.1016/0001-4575(95)00026-V
- Bäumler, H. (2002). *Vergleichende Untersuchung von Fußgängerunfällen bei Tag und Nacht unter Berücksichtigung der Reaktionszeitproblematik bei Dunkelheit*. Dissertation, Technische Universität Dresden.
- Burckhardt, M. (1980). Zur Analyse und Synthese von Reaktionszeiten. *Der Verkehrsunfall*, 18(7/8), 161-168.
- Burckhardt, M. (1985). *Reaktionszeiten bei Notbremsvorgängen*. Köln: Verlag TÜV Rheinland.
- Doisl, C. M. (2004). Entwicklung und Evaluierung aktueller multimodaler Ausgaben für eine Kollisionswarnung im KFZ. Unveröffentlichte Diplomarbeit, Technische Universität München.
- Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, 30, 412-431. (Original erschienen 1869). doi:10.1016/0001-6918(69)90065-1
- Fricke, N. (2009). *Gestaltung von zeit- und sicherheitskritischen Warnungen im Fahrzeug*. Dissertation, Technische Universität Berlin.
- Gratzer, W. (2009). Informationsaufnahme beim Kraftfahrer. In H. Burg, & A. Moser (Eds.), *Handbuch Verkehrsunfallrekonstruktion - Unfallaufnahme, Fahrdynamik, Simulation* (pp. 235-250). Wiesbaden: Vieweg+Teubner Verlag.
- Green, M. (2000). "How long does it take to stop?" Methodological analysis of driver perception-brake times. *Transportation Human Factors*, 2(3), 195-216.
- Huemer, A. K., & Vollrath, M. (2011). Driver secondary tasks in Germany: Using interviews to estimate prevalence. *Accident Analysis and Prevention*, 43(5), 1703-1712. doi: 10.1016/j.aap.2011.03.029
- Huemer, A. K., & Vollrath, M. (2012). *Ablenkung durch fahrfremde Tätigkeiten – Machbarkeitsstudie*. Berichte der Bundesanstalt für Straßenwesen, Mensch und Sicherheit, M 225. Bremerhaven: Wirtschaftsverlag NW.
- Joos, M., Rötting, M. & Velichkovsky, B. M. (2003). Bewegungen des menschlichen Auges: Fakten, Methoden und innovative Anwendungen. In T. Hermann, S. Deutsch & G. Rickheit (Hrsg.), *Handbuch der Psycholinguistik* (S.142-168). Berlin, New York: DeGruyter.
- Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. Report no. DOT HS 810 594*. Washington, D.C.: National Highway Traffic Safety Administration.
- Lamble, D., Kauranen, T., Laakso, M., & Summala, H. (1999). Cognitive load and detection thresholds in car following situations: Safety implications for using

- mobile (cellular) telephones while driving. *Accident Analysis and Prevention*, 31(6), 617-623. doi: 10.1016/S0001-4575(99)00018-4
- Lee, J. D., Ries, M. L., McGehee, D. V., Brown, T. L., & Perel, M. (2000). *Can collision warning systems mitigate distraction due to in-vehicle devices?* Washington, DC: National Highway Traffic Safety Administration.
- Liang, Y., & Lee, J. D. (2010). Combining cognitive and visual distraction: Less than the sum of its parts. *Accident Analysis and Prevention*, 42(3), 881-890. doi: 10.1016/j.aap.2009.05.001
- Löhle, U. (1983). Neue wissenschaftliche Erkenntnisse zur Reaktionsdauer von Kraftfahrern - Konsequenzen für die Verkehrsrechtsprechung und Gutachtenerstellung. *Verkehrsunfall und Fahrzeugtechnik*, 5, 139-143.
- Regan, M. A., Hallett, C., & Gordon, C. P. (2011). Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accident Analysis and Prevention*, 43(5), 1771-1781. doi: 10.1016/j.aap.2011.04.008
- Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science*, 12(6), 462-466. doi: 10.1111/1467-9280.00386
- Summala, H. (2000). Brake reaction times and driver behavior analysis. *Transportation Human Factors*, 2(3), 217-226. doi: 10.1207/STHF0203_2
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97-136. doi:10.1016/0010-0285(80)90005-5
- Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63-102). New York: Academic Press.
- Wickens, C. D., & McCarley, J. S. (2008). *Applied attention theory*. Boca Raton: CRC Press Taylor & Francis Group.
- Zöller, H., & Hugemann, W. (1998, September). *Zur Problematik der Bremsreaktionszeit im Straßenverkehr*. Paper presented at 37. BDP-Kongress Für Verkehrspsychologie, Braunschweig.

Modellierung der Bremsreaktion des Fahrers zur Bewertung von aktiven Fußgängerschutzsystemen - Bestimmung des Bremsverhaltens im Fahrsimulator

Dominik Raudszus¹, Johanna Josten¹, Adrian Zlocki² und Lutz Eckstein¹

Schlüsselwörter: Fahrermodell, Bremsverhalten, aktive Sicherheit, Fußgängerschutz, Simulatorstudie

Zusammenfassung

Es gibt Bestrebungen, die Fahrerreaktion in Testverfahren zur Bewertung der Effektivität von aktiven Fußgängerschutzsystemen zu berücksichtigen. Die Untersuchung der Bremsreaktion des Fahrers auf einen Fußgänger, der plötzlich die Straße betritt, sowie auf entsprechende Warnsignale ist Inhalt dieses Beitrags. Dazu wurde zunächst eine Literaturrecherche durchgeführt, um geeignete Ansätze zur Modellierung des Bremsverhaltens sowie Einflussparameter zu identifizieren. Davon ausgehend wurden die zwei Einflussparameter Kritikalität der Situation und Verfügbarkeit einer Warnung als unabhängige Variablen definiert, während sich die abhängigen Variablen aus festgelegten Modellparametern ergeben. Im Rahmen einer Fahrsimulatorstudie wurde der Einfluss der unabhängigen Variablen auf ausgewählte Parameter der Fahrerreaktion überprüft. Dabei zeigte sich, dass eine Warnung einen signifikanten Einfluss auf die Reaktionszeit des Fahrers hat. Anhand der Ergebnisse der Studie konnte ein entsprechendes Modell parametrisiert werden.

Motivation

Aufgrund der zunehmenden Verbreitung von aktiven Sicherheitssystemen im Kraftfahrzeug ist zur Bewertung der Fahrzeugsicherheit eine ausschließliche Betrachtung der passiven Sicherheitseinrichtungen, wie sie in konventionellen Crash-tests betrieben wird, nicht mehr ausreichend (van Ratingen et al., 2011). Stattdessen muss die Wechselwirkung zwischen aktiven und passiven Systemen betrachtet werden. Aus diesem Grund werden Testverfahren entwickelt, die die Bewertung von aktiven Fußgängerschutzsystemen erlauben. Da solche Systeme im Regelkreis Fahrer - Fahrzeug - Fahrumgebung wirken, kann eine sinnvolle Systembewertung nicht erreicht werden, ohne die Fahrerreaktion zu berücksichtigen. Die Modellierung des Bremsverhaltens als Reaktion auf plötzlich auftauchende Fußgänger unter Berücksichtigung von Systemwarnungen wird im Folgenden beschrieben. Dabei wird ein besonderer Fokus auf die Bestimmung der Fahrerreaktion mit Hilfe einer Fahrsimulatorstudie gelegt.

¹ Institut für Kraftfahrzeuge, RWTH Aachen University

² Forschungsgesellschaft Kraftfahrwesen mbH Aachen

Bewertung aktiver Fußgängerschutzsysteme

Fahrerassistenzsysteme stehen in Wechselwirkung mit dem Fahrzeug, der Umgebung und dem Fahrer. Diese drei Systemelemente interagieren wiederum untereinander. Eine umfassende Bewertung aktiver Fußgängerschutzsysteme muss also diese Elemente ebenfalls berücksichtigen. Testverfahren werden derzeit in unterschiedlichen Projekten entwickelt, wie beispielsweise dem EU-Projekt AsPeCSS (Wisch et al., 2013), der durch deutsche OEMs gegründeten Initiative vFSS (Niewöhner et al., 2011) oder von der durch britische Versicherungen getragenen Forschungseinrichtung Thatcham (Leonard et al., 2012). In den Testverfahren kommen Serienfahrzeuge zum Einsatz, die über ein entsprechendes Fußgängerschutz-System verfügen. Relevante Elemente der Umgebung stellen in diesem Fall der Fußgänger und eine eventuelle Sichtverdeckung dar. Der Fußgänger wird im Testverfahren durch einen Dummy abgebildet, der durch ein Antriebssystem zu einem definierten Zeitpunkt in den Pfad des Testfahrzeugs bewegt wird. Auf Basis der durch das System bewirkten Geschwindigkeitsreduktion kann unter Einbeziehung der passiven Sicherheit die Reduktion des Verletzungsrisikos bestimmt werden. Es sind also Fahrzeug, Umgebung und Fahrerassistenzsystem in der beschriebenen Konfiguration berücksichtigt. Zusätzlich muss jedoch das Fahrerverhalten in die Bewertung aufgenommen werden. Dies spielt bei Systemen eine Rolle, die den Fahrer vor einer drohenden Kollision mit einem Fußgänger warnen. Um eine hohe Reproduzierbarkeit zu gewährleisten, wird auf den Einsatz eines Testfahrers verzichtet und die Fahrerreaktion stattdessen mit Hilfe eines Fahrermodells, das die Sollwerte für einen Bremsroboter vorgibt, simuliert.

Methodisches Vorgehen

Im Folgenden wird die Methodik zur Entwicklung eines geeigneten Fahrermodells beschrieben. Ausgehend von der Anforderungsdefinition wurden zunächst im Rahmen einer Literaturrecherche Informationen über existierende Fahrermodelle zusammengetragen und in Bezug auf die Anwendung im aktiven Fußgängerschutz bewertet. Dabei wurden sowohl deskriptive, als auch funktionale Modelle berücksichtigt. Ein Überblick über unterschiedliche Modellansätze ist beispielsweise bei Michon (1985) oder Ranney (1994) zu finden.

Funktionale Modelle beschreiben die kognitiven Zusammenhänge, die eine Aktion des Fahrers bewirken. Sie erlauben also eine Modellierung der Ursachen des Handelns, während deskriptive Modelle lediglich einen - meist mathematischen - Zusammenhang zwischen Eingangsgrößen und Ausgangsgrößen herstellen und den Fahrer als Blackbox behandeln.

Die Fahrermodelle wurden mit Bezug auf die Anforderungen, die sich aus dem oben beschriebenen Testverfahren ergeben, bewertet. Das Modell soll zum einen in der Lage sein, unterschiedliche Einflussfaktoren zu berücksichtigen und als Modellausgang einen Bremspedalkraftverlauf auszugeben, der sich mit Hilfe eines Bremsroboters umsetzen lässt. Zum anderen soll das Modell über möglichst weni-

ge Parameter verfügen, um die Validierung zu erleichtern und ein transparentes Bewertungsverfahren zu ermöglichen. Aus diesem Grund sind funktionale Modelle für die beabsichtigte Anwendung ungeeignet, da zahlreiche Parameter zur Beschreibung kognitiver Prozesse vorhanden sind, die bei der Bewertung von Fußgängerschutzsystemen nicht von Interesse sind. Es wurde daher mit dem Crossover-Modell nach McRuer et al. (1965) ein deskriptives Modell ausgewählt, das lediglich über zwei Parameter verfügt. Das Open-Loop-Übertragungsverhalten des Fahrers wird dabei als Kombination eines Totzeitelements, das die Reaktionszeit des Fahrers repräsentiert, einem proportionalen Faktor Gain und einem integrierenden Element dargestellt. Als Regelgröße wurde die Bremspedalkraft gewählt. Für die im Folgenden beschriebenen Untersuchungen wurde außerdem die Sollgröße, in diesem Fall die maximale erreichte Bremspedalkraft F_{max} , als dritter Modellparameter berücksichtigt.

Des Weiteren wurden Faktoren identifiziert, die das Bremsverhalten des Fahrers beeinflussen. Zahlreiche Veröffentlichungen (Chang et al., 2008; Forkenbrock et al., 2011; Scott & Gray, 2008) geben an, dass die Fahrerreaktion durch eine Warnung beeinflusst wird. Wogalter (2002) vertritt außerdem die Ansicht, dass auch die Fahrereigenschaften wie z.B. Alter, Erfahrung oder Fahrertyp Einfluss haben. Laut Summala (2000) hängt die Fahrerreaktion auch von der Kritikalität der Situation ab. Ein weiterer Parameter, der die Fahrerreaktion beeinflusst, ist laut Lee et al. (2002) die Ablenkung des Fahrers.

Unter Berücksichtigung der Modellparameter und der Einflussfaktoren ließ sich eine Abhängigkeitsmatrix erstellen, die alle möglichen Zusammenhänge zwischen Einflussfaktoren und Parametern enthält. Zunächst konnten einige Informationen der Literatur entnommen werden. Dabei zeigte sich allerdings, dass einige Studien zu widersprüchlichen Ergebnissen kommen. Zudem gelten viele Ergebnisse nur für bestimmte Randbedingungen und sind somit nur bedingt übertragbar. Für einige Kombinationen sind außerdem keine Informationen verfügbar. Aus diesem Grund wurde eine Fahrsimulatorstudie durchgeführt, um den Einfluss ausgewählter Faktoren auf die Modellparameter zu bestimmen.

Aus den Informationen, die in der Abhängigkeitsmatrix erfasst wurden, lassen sich die folgenden Hypothesen für die Probandenstudie ableiten:

- Eine Warnung reduziert die Reaktionszeit,
- Eine Warnung erhöht den Faktor Gain,
- Eine Warnung erhöht die maximale Bremspedalkraft,
- Eine hohe Kritikalität reduziert die Reaktionszeit,
- Eine hohe Kritikalität erhöht den Faktor Gain,
- Eine hohe Kritikalität erhöht die maximale Bremspedalkraft.

Versuchsdesign

Wie oben beschrieben, soll die Abhängigkeit der Modellparameter Gain, Fmax und Reaktionszeit von den wichtigsten Einflussfaktoren bestimmt werden. Aus diesem Grund wird eine Fahrsimulatorstudie durchgeführt, deren Versuchsdesign in den folgenden Abschnitten beschrieben wird.

Festlegen der Testgruppen

Um den Umfang der Studie zu begrenzen, wurde zunächst die Anzahl der zu untersuchenden Einflussfaktoren reduziert. Von besonderer Bedeutung für das oben beschriebene Testverfahren ist der Einfluss einer Warnung auf das Fahrerverhalten, verglichen mit dem Fahrerverhalten ohne Warnung. Des Weiteren ist der Einfluss der Kritikalität der Situation besonders relevant, da diese in den oben beschriebenen Testszenarien ebenfalls variiert. Dabei wird die Kritikalität durch die Time to Collision (TTC), die den Quotienten von Abstand zwischen Pkw und Fußgänger und der Fahrzeuggeschwindigkeit angibt, repräsentiert. Eine geringe TTC entspricht somit einer hohen Kritikalität. Der TTC-Wert bezieht sich auf den Zeitpunkt, an dem der Fußgänger für den Fahrer sichtbar wird. Für die Fahrsimulatorstudie ergeben sich also die unabhängigen Variablen Warnung und TTC, sowie die abhängigen Variablen Gain, Fmax und Reaktionszeit. Für die TTC wurden zwei unterschiedliche Werte untersucht.

Das Festlegen der TTC zur Erzeugung einer ausreichend kritischen Situation erfolgte unter der Annahme, dass eine Verzögerung von 1 g umgesetzt wird. Bei einer Geschwindigkeit von 50 km/h muss somit bei einer TTC von 0,7 s gebremst werden, um eine Kollision mit dem Fußgänger zu vermeiden. Auf Basis von anderen Studien (Green, 2000; Olson & Sivak, 1986) wird eine Reaktionszeit von etwa 1 s bis 1,2 s erwartet. Daraus ergibt sich, dass der Fahrer eine Kollision gerade noch vermeiden kann, wenn der Fußgänger bei einer TTC von 1,8 s sichtbar wird. Außerdem wird eine weniger kritische Situation getestet. Die Differenz sollte einen spürbaren Unterschied in der subjektiv wahrgenommenen Kritikalität bewirken. In Voruntersuchungen stellte sich ein Wert von 2,5 Sekunden als geeignet heraus. Aus den unabhängigen Variablen resultieren also vier Testgruppen, die im Folgenden zusammengefasst sind:

- Warnung, TTC = 1,8 s
- Keine Warnung, TTC = 1,8 s
- Warnung, TTC = 2,5 s
- Keine Warnung, TTC = 2,5 s

Testszenario

Aus der Unfallforschung geht hervor, dass Szenarien, in denen der Fußgänger vor einem Pkw, der geradeaus fährt, die Straße überquert, am häufigsten auftreten

(Wisch et al., 2013). Um eine definierte TTC umsetzen zu können, wird für die Fahrsimulatorstudie ein Szenario gewählt, in dem der Fußgänger hinter einer Verdeckung hervortritt.

Die Probandenstudie wurde im Fahrsimulator des ika durchgeführt. Dieser verfügt über ein Bewegungssystem mit sechs Freiheitsgraden sowie über ein Sichtsystem mit 180° Projektion und kann mit unterschiedlichen Fahrzeug-Mock-Ups betrieben werden. Während des beschriebenen Versuchsaufbaus kam ein Mittelklasse-Pkw mit Automatik-Getriebe und Head-Up Display zum Einsatz.

Abb. 2 zeigt das implementierte Testszenario. Auf der rechten Seite sind parkende Fahrzeuge zu sehen, die genutzt werden, um den Fußgänger zu verdecken. Der Fußgänger erscheint hinter den Fahrzeugen und überquert mit konstanter Geschwindigkeit die Straße. Um die Belastung der Probanden zu reduzieren, setzt der Fußgänger auch im Falle einer Kollision seine Bewegung fort. Dadurch entsteht der subjektive Eindruck, dass der Fußgänger nicht getroffen wurde. Die gesamte Fahrt findet in städtischer Umgebung statt, während regelmäßig Umgebungsverkehr und nicht-querende Fußgänger auftauchen. Bis zum Sichtbarwerden des Fußgängers betrug die Fahrdauer etwa 5 Minuten. Die Probanden waren angewiesen eine Geschwindigkeit von 50 km/h einzuhalten.

Die Fußgängerwarnung wurde in den entsprechenden Testgruppen ebenfalls zu dem Zeitpunkt, an dem der Fußgänger für den Fahrer sichtbar wird, ausgegeben. Dabei wurde ein akustisches Signal (Piepton) mit einem Warnsymbol (Warndreieck mit Fußgängersymbol) im Head-Up Display kombiniert.



Abb. 2: Darstellung des Testszenarios im Fahrsimulator

Um eine repräsentative Beanspruchung durch die Fahraufgabe zu erreichen und eine plausible Cover-Story zu gewährleisten, hatten die Probanden eine Sekundäraufgabe zu absolvieren, die darin bestand, den Anweisungen eines laut Instruktion neuartigen Navigationssystems zu folgen. Diese Anweisungen wurden ausschließlich akustisch ausgegeben, sodass der Blick des Fahrers nicht von der Straße abgelenkt wurde. Eine erste Navigationsansage erfolgte etwa 300 Meter vor jeder Kreuzung. Kurz bevor der Fahrer die Kreuzung erreichte, erhielt er einen weiteren akustischen Navigationshinweis. Der letzte Hinweis mit der Anweisung "Gerade-

aus fahren" wurde etwa 15 Sekunden vor dem Auftauchen des Fußgängers ausgegeben.

Versuchsablauf

Um Erwartungseffekte, die signifikanten Einfluss auf die Fahrerreaktionszeit haben, wie im FP7-Projekt ASSESS (Ranovona et al., 2011) gezeigt wurde, auszuschließen, wurde das Szenario nur einmal pro Proband durchgeführt. Außerdem erhielten die Probanden vorab keine Informationen zum tatsächlichen Ziel der Studie, sondern es wurde entsprechend der Cover-Story vorgegeben, dass ein Navigationssystem untersucht werden soll.

Vor dem Beginn des eigentlichen Testszenarios absolvierte jeder Proband ein Eingewöhnungsszenario, in dem der Versuchsleiter die grundlegenden Funktionen des Fahrzeug-Mock-Ups sowie alle weiteren Versuchseinrichtungen erklärte. Während dieses Szenarios fuhr der Proband in urbaner Umgebung, die dem späteren Testszenario bereits sehr ähnlich war. Es wurden jedoch keine Hinweise des Navigationssystems ausgegeben. Um einen Eindruck von der Bremscharakteristik zu vermitteln, wurden die Probanden während der Gewöhnungsfahrt unter dem Vorwand eines besseren Kennlernens des Fahrzeugs gebeten einige Bremsmanöver durchzuführen.

Während des Testszenarios wurden Bewegungsdaten und Fahrereingaben mit einer Abtastzeit von 10 ms aufgezeichnet. Gleichzeitig wurde ein Flag, das angibt, wann der Fußgänger startet, aufgenommen.

Sowohl vor, als auch nach der Fahrt wurden die Probanden gebeten Fragebögen auszufüllen. Erfasst wurden unter anderem Angaben zur Demographie und dem Fahrverhalten sowie in der Nachbefragung eine Einschätzung der Situation hinsichtlich der Notwendigkeit einer Bremsung und der wahrgenommenen Kritikalität.

Im Anschluss an den Versuch wurden die Modellparameter Reaktionszeit, Gain und maximale Bremskraft für jeden Probanden bestimmt, indem die Parameter iterativ so angepasst wurden, dass die Modell-Ausgabe und der gemessene Pedalkraftverlauf größtmöglich übereinstimmten. Als Referenzzeitpunkt ($t = 0$) wurde der Zeitpunkt genutzt, an dem der Fußgänger für den Fahrer sichtbar wird. Zur selben Zeit wurde auch das Warnsignal ausgegeben. Die auf diese Weise bestimmte Reaktionszeit gibt also das Zeitintervall zwischen erster Sichtbarkeit des Fußgängers und Beginn der Pedalbetätigung an.

Versuchsergebnisse

Im Rahmen der Fahrsimulatorstudie wurden 27 Versuchspersonen getestet. Die Datensätze von zwei Versuchspersonen konnten aufgrund fehlerhafter Messaufzeichnungen nicht in die Analyse einbezogen werden. Alle Auswertungsdaten beziehen sich daher auf die verbleibenden 25 Personen mit einem Durchschnittsalter

von 31 Jahren [Standardabweichung (SD) 9,8 Jahre, davon 23 männlich]. Die durchschnittliche Fahrerfahrung der Stichprobe betrug 13 Jahre (Minimum 5 Jahre, SD 9,4 Jahre) bei einer durchschnittlichen jährlichen Fahrleistung von 13.310 km (Minimum 250 km, SD 8194 km). Über 50 Prozent der Teilnehmer nutzen das Auto mindestens viermal in der Woche.

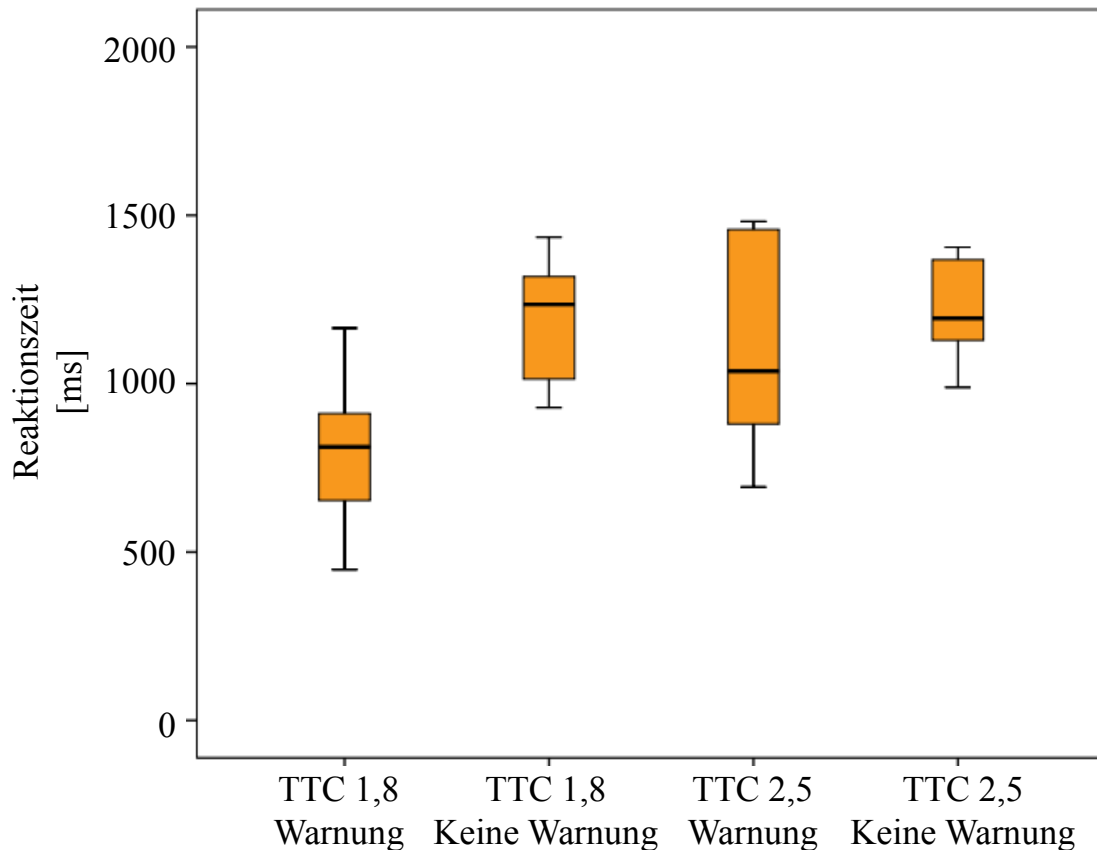


Abb. 3: Boxplots des Parameters Reaktionszeit für alle Testgruppen

Abb. 3 zeigt einen Box-Plot der Reaktionszeit für die vier Versuchsgruppen. Beim Vergleich von Gruppe 1 (mit Warnung, TTC 1,8 sec) und Gruppe 2 (Keine Warnung, TTC 1,8 sec) wird deutlich, dass die Warnung einen großen Einfluss auf die Reaktionszeit hat, wenn die TTC gering ist. Bei höherer TTC hat die Warnung einen abnehmenden Einfluss. Bei Betrachtung der Gruppen 3 und 4 wird jedoch auch deutlich, dass die Streuung der Reaktionszeit deutlich größer ist, wenn eine Warnung ausgegeben wird. Dies könnte dadurch verursacht sein, dass einige Probanden durch die frühe Warnung irritiert sind, während andere Probanden von der Warnung profitieren. Insgesamt lässt sich jedoch erkennen, dass der Median der Reaktionszeit deutlich geringer ist, wenn eine Warnung ausgegeben wird, als wenn keine Warnung ausgegeben wird. Die statistische Auswertung der Reaktionszeiten mittels ANOVA ergibt einen signifikanten Haupteffekt der Warnung [$F(1,21) = 5,482$, $p < .05$] mit geringeren Reaktionszeiten für vorgewarnte Fahrer (mit Warnung 960 ms versus 1200 ms ohne Warnung). Der Haupteffekt der Kritikalität (kurze TTC versus lange TTC) ist lediglich tendenziell signifikant [$F(1,21) = 3,106$, $p < .1$] mit höheren Reaktionszeiten bei einer langen TTC (990 ms versus

1170 ms). Die Interaktion zwischen den beiden Variablen ist jedoch nicht signifikant [$F(1,21) = 2,247$, $p = .149$].

Der Mittelwert der maximalen Bremspedalkraft über alle Testgruppen beträgt 290,15 N bei einer Standardabweichung von 70,69 N. Die Analyse für die maximale Bremskraft ergibt keinen signifikanten Effekt der Warnung [$F(1,21) = 1,71$, $p = .205$; mit Warnung (250 N) versus ohne Warnung (291 N)] oder der TTC [$F(1,21) = 0,003$, $p = .954$; kurze TTC (272 N) versus lange TTC (270 N)]. Die Interaktion zwischen diesen beiden Parametern ist ebenfalls nicht signifikant [$F(1,21) = 0,455$, $p = .507$].

Für den Parameter Gain wurde ein Mittelwert von 3,05 s⁻¹ und eine Standardabweichung von 1,39 s⁻¹ bestimmt. Aufgrund der Verteilung des Parameters Gain wurde ein nicht-parametrisches Testverfahren angewandt. Die Daten eines Probanden wurden als Ausreißer gewertet und von der Analyse ausgeschlossen, da der z-transformierte Wert für den Parameter Gain den vorab definierten kritischen Wert von 3,29 z überschreitet. Damit gehen 24 Personen in die Analyse mit ein. Es wurde kein signifikanter Unterschied zwischen den Gruppen gefunden [$\chi^2(3) = 5,783$, $p = .123$].

Die Produkt-Moment-Korrelation nach Pearson zwischen Fmax und dem Modellparameter Reaktionszeit zeigt tendenzielle Signifikanz [$r = .388$, $p = .055$]. Dies deutet darauf hin, dass eine schnellere Fahrerreaktion eine geringere maximale Bremspedalkraft bewirkt, und umgekehrt, dass ein Fahrer, der spät bremst, eine höhere Bremspedalkraft aufbringt. Reaktionszeit und Gain [$r = -.008$, $p = .972$] sowie Gain und Fmax [$r = .217$, $p = .308$] korrelieren nicht miteinander.

Alle Teilnehmer, die eine Warnung erhielten, nahmen diese auch wahr und konnten diese, bis auf eine Ausnahme, anschließend zutreffend beschreiben. Nur ein Proband gab an, keine unbedingte Notwendigkeit zur Bremsung in der kritischen Fahrsituation gesehen zu haben. Die Auswertung der subjektiven Einschätzung ergab jedoch wenige Unterschiede in der Beurteilung der Situation zwischen den Gruppen. Beispielsweise veränderte sich die empfundene Kritikalität der Situation weder durch die Variation von Warnung [mit (2,57; SD 1,8) versus ohne Warnung (2,63; SD 1,8)] noch Kritikalität [kurze TCC (2,08; SD 1,5) versus lange TTC (3,07; SD 1,9)]. Erhoben wurde die Einschätzung mittels eines siebenstufigen semantischen Differentials von kritisch zu unkritisch.

Fazit

Es lässt sich zusammenfassen, dass die Verfügbarkeit einer Warnung und tendenziell die Kritikalität der Situation Einfluss auf die Reaktionszeit des Fahrers haben. Somit hat sich die erste Hypothese bestätigt, während die vierte Hypothese detailliertere Untersuchungen erfordert. So könnte beispielsweise eine stärkere Variation der TTC aussagekräftigere Ergebnisse hervorbringen. Ein Einfluss auf die anderen Modellparameter Gain und maximale Bremskraft konnte nicht nachgewie-

sen werden. Die übrigen Hypothesen können daher in dieser Studie nicht bestätigt werden. Für das oben beschriebene Modell bedeutet dies, dass lediglich die Reaktionszeit von der Warnung und der Kritikalität abhängt. Für die anderen Parameter Gain und Fmax können konstante Werte festgelegt werden.

Literatur

- Chang, S.-H., Lin, C.-Y., Hsu, C.-C., Fung, C.-P., & Hwang, J.-R. (2009). The effect of a collision warning system on the driving performance of young drivers at intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(5), 371-380. doi:10.1016/j.trf.2009.05.001
- Forkenbrock, G., Snyder, A., Heitz, M., Hoover, R. L., O'Harra, B., Vasko, S., & Smith, L. (2011). A Test Track Protocol for Assessing Forward Collision Warning Driver-Vehicle Interface Effectiveness. NHTSA Technical Report DOT HS 811 501
- Green, M. (2000). "How Long Does It Take to Stop?" Methodological Analysis of Driver Perception-Brake Times. *Transportation Human Factors*, 2(3), 195-216
- Lee, J. D., McGehee, D. V., Brown, T. L., & Reyes, M. L. (2002). Driver Distraction, Warning Algorithm Parameters, and Driver Response to Imminent Rear-end Collisions in a High-Fidelity Driving Simulator. NHTSA Technical Report DOT HS 809 448
- Leonard, J., Danton, R., Avery, M., Weekes, A., Zuby, D., & Kühn, M. (2012). Typical Pedestrian Accident Scenarios for the Testing of Autonomous Emergency Braking Systems. *Proceedings of the 22nd International Technical Conference on the Enhanced Safety of Vehicles (ESV)*
- McRuer, D. T., Graham, D., Krendel, E. S., & Reisner, W. (1965). *Human Pilot Dynamics in Compensatory Systems*. Franklin Institute, Technical Report
- Michon, J. A. (1985). A Critical View of Driver Behavior Models: What Do We Know, What Should We Do? Human behavior and traffic safety, Springer US, 485-520, doi:10.1007/978-1-4613-2173-6_19
- Niewöhner, W., Roth, F., Gwehenberger, J., Gruber, C., Kuehn, M., Sferco, R., Stanzel, M. (2011). Proposal for a Test Procedure of Assistance Systems regarding Preventive Pedestrian Protection. *Proceedings of the 22nd International Technical Conference on the Enhanced Safety of Vehicles (ESV)*
- Olson, P. L., & Sivak, M. (1986). Perception-Response Time to Unexpected Roadway Hazards. *Human Factors*, 28(1), 91-96.
- Ranney, T. A. (1994). Models of Driving Behavior: A Review of their Evolution. *Accident Analysis and Prevention*, 26(6), 733-750
- Ranovona, M., Tsuchida, J., Chin, E., Unselt, T., Munoz, O., Aparicio, A., & Winkler, B. (2011). Report on the experimental study results of the evaluation of behavioural aspects. ASSESS Deliverable 3.2

- Scott, J. J., & Gray, R. (2008). A Comparison of Tactile, Visual, and Auditory Warnings for Rear-End Collision Prevention in Simulated Driving. *Human Factors*, 50(2), 264-275. doi:10.1518/001872008X250674
- Summala, H. (2000). Brake Reaction Times and Driver Behavior Analysis. *Transportation Human Factors*, 2(3), 217-226
- van Ratingen, M., Williams, A., Castaing, P., Lie, A., Frost, B., Sandner, V.,...Weimer, C. (2011). Beyond NCAP: Promoting New Advancements in Safety. *Proceedings of the 21st International Technical Conference on the Enhanced Safety of Vehicles (ESV)*
- Wisch, M., Seiniger, P., Pastor, C., Edwards, M., Visvikis, C., & Reeves, C. (2013). Scenarios and weighting factors for pre-crash assessment of integrated pedestrian safety systems. AsPeCSS Deliverable 1.1
- Wogalter, M. S., Conzola, V. C., & Smith-Jackson, T. L. (2002). Research-Based Guidelines for Warning Design and Evaluation. *Applied Ergonomics*, 33, 219-230

Fahrerseitiger Unterstützungsbedarf in drohenden Verkehrskonfliktszenarien und Wirksamkeitsuntersuchung frühzeitiger Fahrerinformationen basierend auf kooperativer Umfelderkennung

Frederik Naujoks, Heidi Grattenthaler und Alexandra Neukum¹

Schlüsselwörter: Kooperative Umfelderkennung, Fahrerinformationen, Fahrsimulation, Unterstützungsbedarf.

Zusammenfassung

Kooperative Umfelderkennung ermöglicht – über Notfallwarnungen hinaus – frühzeitige Informationen über potenzielle Gefahrensituationen an den Fahrer zu übermitteln. In Anlehnung an Forschungsergebnisse zu Fahrerwarnungen wurde erwartet, dass die Effektivität solcher Fahrerinformationen von situationsspezifischen Fahrererwartungen abhängt. In einer Simulatorstudie durchführten N = 20 Probanden eine breite Auswahl an Längsverkehrs- und Kreuzungsszenarien von unterschiedlicher Vorhersehbarkeit der Situationsentwicklung. Die Szenarien wurden in verschiedenen Sichtbedingungen mit und ohne eine auf kooperativer Umfelderkennung basierende Fahrerunterstützung absolviert (visuell-auditive Fahrerinformationen über bevorstehende Konflikte zwei Sekunden vor dem letzten Warnzeitpunkt). Die Ergebnisse zeigen eine deutliche Situationsentschärfung durch frühzeitige Fahrerinformationen in allen erhobenen Konfliktszenarien mit fahrerseitigem Unterstützungsbedarf, besonders in überraschenden und unerwarteten Szenarien, und verdeutlichen somit das Potenzial kooperativer Umfelderkennung.

Einleitung: Fahrerunterstützung bei Verkehrskonflikten

Das Forschungsprojekt Ko-PER zielt mittels einer möglichst vollständigen Erfassung der lokalen Verkehrsumgebung auf Verbesserungen der präventiven Sicherheit im Straßenverkehr ab. Über Fahrzeug-Fahrzeug- und Fahrzeug-Infrastruktur-Kommunikation werden Informationen über das Verkehrsgeschehen gewonnen und mit fahrzeuglokaler Umfeld-Erfassung fusioniert (sog. kooperative Perzeption). Somit können potentielle Konfliktsituationen vollständiger und rechtzeitiger erkannt werden als bisher, was eine frühzeitige Fahrerunterstützung ermöglicht.

Die Arbeiten zur Gestaltung der Mensch-Maschine-Schnittstelle verfolgen im Projekt Ko-PER das Ziel, geeignete Informationsstrategien zur Fahrerunterstützung in zeitkritischen Konfliktsituationen zu ermitteln. Etablierte, auf fahrzeuglokaler Sensorik basierende Konzepte zur Fahrerunterstützung in Konfliktsituationen se-

¹ Interdisziplinäres Zentrum für Verkehrswissenschaften an der Universität Würzburg (IZVW), Röntgenring 11, 97070 Würzburg, mail: naujoks@psychologie.uni-wuerzburg.de

hen vor, Kollisionen mittels Darbietung dringlicher Fahrerwarnungen zu vermeiden und ggf. autonom in die Fahrzeugführung einzugreifen (z.B. Brown, Lee & McGehee, 2001; Rhede, Wäller & Oel, 2011; Winner, 2009). Ziel hierbei ist es, den Fahrer zu einer unmittelbaren Brems- oder Lenkreaktion zu bringen (sog. „imminent crash warnings“, Lenné & Triggs, 2009).

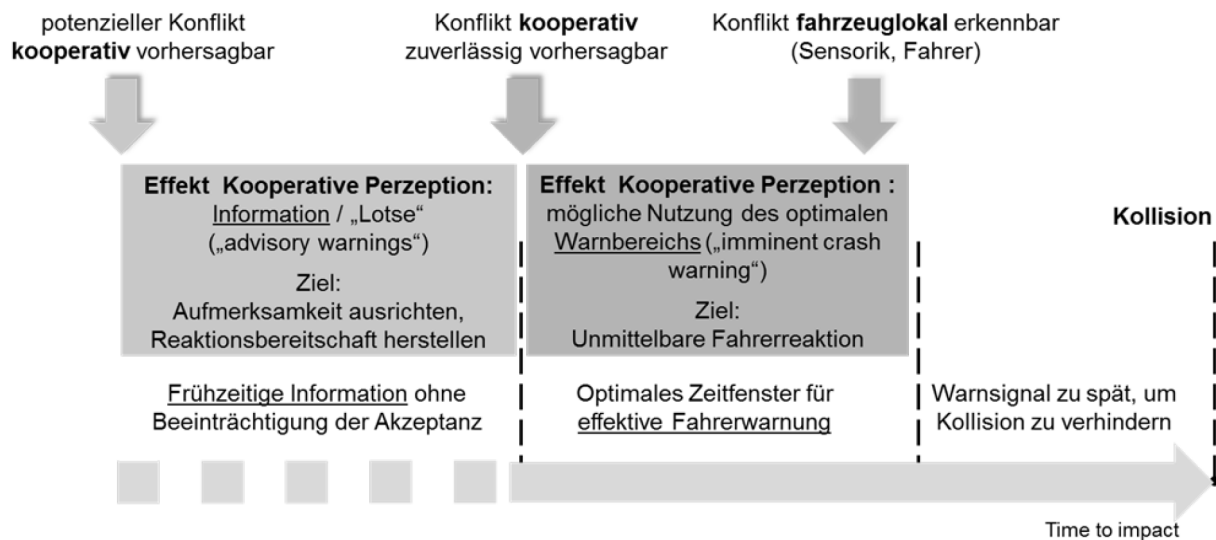


Abb.1: Schematische Darstellung der Zeitfenster zur Fahrerunterstützung in den letzten Sekunden vor einer Kollision (Neukum, 2011).

Durch kooperative Perception wird das Zeitfenster vor dringlichen Fahrerwarnungen für die Darbietung von sog. „Fahrerinformationen“ zugänglich (Neukum, 2011, siehe Abb.1). Diese auch als „advisory warnings“ (Lenné & Triggs, 2009) oder „Risikoinformationen“ (Rhede et al., 2011) bezeichneten Warnsignale sollen die Aufmerksamkeit des Fahrers auf einen potenziellen Konflikt richten und Reaktionsbereitschaft herstellen (Neukum, 2011). Die zeitliche Struktur von Warnkonzepten, die auf fahrzeuglokaler Sensorik basieren, wird somit mittels kooperativer Sensorik um eine frühzeitige Informationskomponente ergänzt.

Gestaltungsaspekte der Fahrerunterstützung

Grundlage für die Ausgabe eines Informations- oder Warnsignals auf Basis kooperativer Sensorik stellt die sog. Situationsanalyse dar (Weidl & Breuel, 2012). Mittels fusionierter Daten der kooperativen Perception wird eine probabilistische Situationsprognose erstellt und ggf. eine Fahrerinformation oder –warnung ausgegeben. Dies stellt eigene Anforderungen an die Gestaltung der Mensch-Maschine-Schnittstelle, da sich die zu erwartende Unzuverlässigkeit der Situationsprognose negativ auf Effektivität und Akzeptanz der Fahrerinformationen auswirken könnte (Bliss & Acton, 2003; Sorkin, 1988). Hinsichtlich der Gestaltung der Mensch-Maschine-Schnittstelle sind folgende Punkte hervorzuheben:

(1) Zeitliche Gestaltung: Informationen sollten so spät wie möglich präsentiert werden, allerdings so früh wie nötig, um noch eine adäquate Fahrerreaktion einleiten zu können. Naujoks, Grattenthaler und Neukum (2012) konnten zeigen, dass visuell-auditive Fahrerinformationen 1-2 Sekunden vor dem letztmöglichen Warnzeitpunkt zu einer deutlichen Entschärfung von Verkehrskonflikten gegenüber dem nicht-assistierten Fahren führten.

(2) Modalität: Gestaltungsempfehlungen zur Modalität von Fahrerinformationen lassen sich wie folgt zusammenfassen (z.B. Campbell, Carney & Kantowitz, 1997; COMSIS Corporation, 1996; Dingus, Jahns, Horowitz & Knipling, 1998; Green, Levison, Paelke & Serafin, 1993; International Harmonized Research Activities (IHRA) working group on Intelligent Transport Systems (ITS), 2008; Rhede et al., 2011): Verwendung visueller Anzeigen, anstelle von Sprachmeldungen oder dringlichen Tönen. Ein ankündigender, nicht aufdringlicher Ton kann die Effektivität der Information hingegen steigern.

(3) Informationsinhalt: Es kann empfohlen werden, zusätzlich die Art des jeweiligen Konflikts und den Ort, an dem der Konflikt droht, darzustellen.

Fahrerseitiger Unterstützungsbedarf in Verkehrssituationen

Die Fahrzeugführung kann als dynamische Kontrollaufgabe gesehen werden: Fahrer überwachen Veränderungen der Verkehrsumgebung kontinuierlich, um Kollisionen zu vermeiden (Koorstra, 1993; Van der Hulst, Meijman & Rothengatter, 1999). Aufgrund der Informationen aus der Verkehrssituation und dem Verhalten anderer Verkehrsteilnehmer werden Veränderungen der Verkehrsumgebung abgeschätzt, um das eigene Verhalten an zukünftige Bedingungen anpassen zu können. Fahrer schätzen somit kontinuierlich das Kollisionspotential ab und lenken entsprechend ihre Aufmerksamkeit (Brown, Lee & McGehee, 2000). So werden beispielsweise Objekte wie Verkehrszeichen oder andere Verkehrsteilnehmer in Verkehrsszenen schneller an erwarteten als an unerwarteten Orten entdeckt (Theeuwes & Hagenzieker, 1993). Die Erwartung, dass ein Abbremsen in naher Zukunft nötig sein wird, führt entsprechend zu schnelleren Fahrerreaktionen (Muhler & Vollrath, 2010; Schweitzer et al., 1995). In überraschenden Situationen zeigen Fahrerwarnungen folglich die größte Wirksamkeit (Schmidt & Krüger, 2010). Es wird dementsprechend eine Abhängigkeit der Versuchsergebnisse von der Möglichkeit zur Ausbildung von Fahrererwartungen über einen bevorstehenden Konflikt in der jeweiligen Prüfsituation angenommen.

Fragestellung

Vorgestellt werden nachfolgend Ergebnisse einer Simulatoruntersuchung zu den Fragestellungen: (1) In welchen Fahrsituationen besteht ein fahrerseitiger Unterstützungsbedarf? Um dies zu beantworten, wurden Fahrer mit unterschiedlichsten Konfliktsituationen konfrontiert, wobei die Sichtbedingungen und die Möglichkeit zur Antizipation von Verkehrskonflikten variiert wurden. (2) Können Konfliktsitu-

ationen durch frühzeitige Fahrerinformationen entschärft werden? Zur Beantwortung dieser Fragestellung wurde ein prototypisches Unterstützungskonzept erarbeitet und dessen Wirksamkeit überprüft.

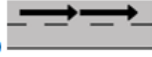
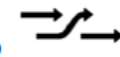
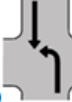










Methodisches Vorgehen

Stichprobe und Versuchsdesign

Stichprobe: Die Studie wurde im statischen Fahrsimulator der WIVW GmbH durchgeführt. Untersucht wurde eine Stichprobe von $N = 20$ Probanden, jeweils zur Hälfte Männer und Frauen und je zur Hälfte jüngere (25 bis 40 Jahre; $M = 30.10$ Jahre, $SD = 4.58$) und ältere Fahrer (ab 57 Jahre; $M = 66.20$ Jahre, $SD = 6.09$). Die Probanden wurden dem Testfahrerpanel der WIVW GmbH entnommen und hatten bereits an einem ausführlichen Simulatortraining teilgenommen.

Szenarien: Mittels einer Anforderungsanalyse wurden Anwendungsfälle, d.h. Gruppen von Fahraufgaben, bei denen der Fahrer basierend auf Ko-PER Kommunikationstechnologie unterstützt werden soll, erarbeitet und insgesamt 13 unterschiedliche Prüfzenarien definiert (siehe Tab.1).

Tab.1: Anwendungsfälle und Szenarien (schematische Darstellungen aus GDV, 1998).

Anwendungsfall	Szenario	Schematische Darstellung
1 Abstandhaltung	1. Auffahren in einer Kolonne	(1) 
2 Spurwechsel	2. Spurwechsel bei Überholvorgängen	(2) 
3 Gegenverkehr	3. Linksabbiegen bei Gegenverkehr 4. Vorbeifahren bei Gegenverkehr	(3)  (4) 
4 Querverkehr	5. Kreuzen – vorfahrtsberechtigtes Fahrzeug von links 6. Kreuzen – vorfahrtsberechtigtes Fahrzeug von rechts 7. Linkseinbiegen – vorfahrtsberechtigtes Fahrzeug von links	(5)  (6)  (7) 
5 Fußgänger	8. Überschreiten 9. Rechtsabbiegender PKW – querender Fußgänger 10. Linksabbiegender PKW – querender Fußgänger 11. Fußgänger in Längsrichtung auf der Fahrbahn	(8)  (9)  (10)  (11) 
6 Radfahrer	12. Kreuzender PKW – querender Radfahrer 13. Rechtsabbiegender PKW – querender Radfahrer	(12)  (13) 

Unabhängige Variablen: Die o.g. Szenarien variierten (1) in der Sichtbedingung (Konfliktpartner ist bei der Anfahrt an den gemeinsamen Konfliktpunkt verdeckt vs. sichtbar), (2) der Möglichkeit zur Antizipation von Konflikten (Erwartung hoch vs. niedrig über Vorfahrtsregelung und Erwartbarkeit des Konflikts), und wurden (3) mit und ohne Fahrerunterstützung absolviert (siehe Tab.2). In der Bedingung mit Fahrerunterstützung wurde jeweils zwei Sekunden vor dem letztmöglichen Warnzeitpunkt eine Information über Art und Richtung des jeweiligen Konflikts im simulierten Head-Up Display (HUD) mit einem ankündigenden Ton (500Hz Sinus) dargeboten. Unter Berücksichtigung der Fahrtgeschwindigkeit wird hierbei davon ausgegangen, dass der Fahrer mit einer Reaktionszeit von einer Sekunde und mit einer konstanten Verzögerung von -8m/s^2 zum Stillstand abbremst (siehe Naujoks et al., 2012 für eine ausführliche Darstellung).

Tab.2: Untersuchte Szenarien. IS = Innenstadt; LS = Landstraße; BAB = Bundesautobahn; XK = schieldergeregelter Kreuzung, XKAB = ampelgeregelter Kreuzung; VB = Unterstützter Fahrer hat Vorrang oder ist wartepflichtig (WP); mit = Konfliktpartner aus gleicher Fahrtrichtung oder entgegengesetzter Fahrtrichtung (entgegen).¹²

Szenario	Geometrie	Vorfahrt/ Fahrtrichtung	v [km/h]	Erwartung des Konflikts
Auffahren in einer Kolonne	IS/LS/BAB		50	niedrig
Überholen	BAB		130	hoch
Linksabbiegen bei Gegenverkehr	XK	WP	25	hoch
	XK	VB	50	niedrig
Vorbeifahren bei Gegenverkehr	IS		50	hoch
Kreuzen – vorfahrtsberechtigtes Fahrzeug von rechts	XK	WP	50	hoch
	XK	VB	50	niedrig
Kreuzen – vorfahrtsberechtigtes Fahrzeug von links	XK	WP	50	hoch
	XK	VB	50	niedrig
Linkseinbiegen – vorfahrtsberechtigtes Fahrzeug von links	XK	WP	25	hoch
	XK	VB	50	niedrig
Überschreitender Fußgänger	IS		50	niedrig
Rechtsabbiegender PKW – querender Fußgänger	XKAB		25	hoch
Linksabbiegender PKW – querender Fußgänger	IS		25	hoch
Fußgänger in Längsrichtung auf der Fahrbahn	IS		50	niedrig
	LS		80	niedrig
Kreuzender PKW – querender Radfahrer	XK	WP	50	hoch
	XK	VB	50	niedrig
Rechtsabbiegender PKW – querender Radfahrer	XKAB	gleich	25	hoch
	XKAB	entgegen	25	hoch

¹ Jedes Szenario wurde in zwei Sichtbedingungen mit und ohne Unterstützung durchfahren (N = 88 Situationen).

² Das Szenario ‚Rechtsabbiegender PKW – querender Radfahrer‘ wurde in zwei Richtungsvarianten absolviert.

Abhängige Variablen und Hypothesen

Erfasste Parameter: Es wurden verschiedene Parameter des Fahrverhaltens betrachtet (siehe Tab.3). Nach jeder absolvierten Situation wurden die Probanden zusätzlich zur Situationskritikalität befragt. Der fahrerseitige Unterstützungsbedarf wird abgeleitet über eine nachfolgend dargestellte Integration von Parametern des Fahrverhaltens und subjektiver Situationskritikalität. Im Anschluss an die Simulatorfahrt wurde weiterhin eine Nachbefragung durchgeführt, in welcher Angaben zum Vorkommen im realen Straßenverkehr, Unfallträchtigkeit sowie dem eingeschätzten Sicherheitsgewinn durch die Fahrerunterstützung pro Szenario zu machen waren.

Tab.3: Übersicht abhängige Variablen

Parameter	Beschreibung	Einheit/Skala
Bremsreaktion und zur Verfügung stehende Reaktionszeit zur Kollisionsvermeidung		
Maximale Bremspedalstellung	Maximale Betätigung des Bremspedals beim Abbremsen	[% der maximalen Bremspedalstellung]
Time-to-collision (TTC)	Zeit bis zur Kollision mit Konfliktpartner bei gleicher Bewegungsrichtung	[s]
Time-to-arrival (TTA)	Zeit bis zur Kollision mit Konfliktpartner bei kreuzender Bewegungsrichtung	[s]
Fahrtbefragung		
Situationskritikalität (Beurteilung durch Proband und Versuchsleiter)	Wie kritisch war die Situation?	0-10, Situationsbeurteilungsskala (Neukum et al., 2008)
Nachbefragung		
Vorkommen	Die Situation kommt im Alltag häufig vor.	
Unfallträchtigkeit	In der Situation passieren im Alltag häufig Unfälle.	trifft nicht zu – neutral – trifft zu
Sicherheitsgewinn durch Fahrerunterstützung	Durch die Information wurde das Fahren sicherer.	

Ableitung des Unterstützungsbedarfs: Die Auftretenshäufigkeiten kritischer Begegnungen in den nicht-assistierte Fahrten spiegeln die Kritikalität und gleichzeitig den Unterstützungsbedarf der jeweiligen Situation wider. Zu deren Bestimmung werden objektive und subjektive Daten integriert (vgl. Abb.3).

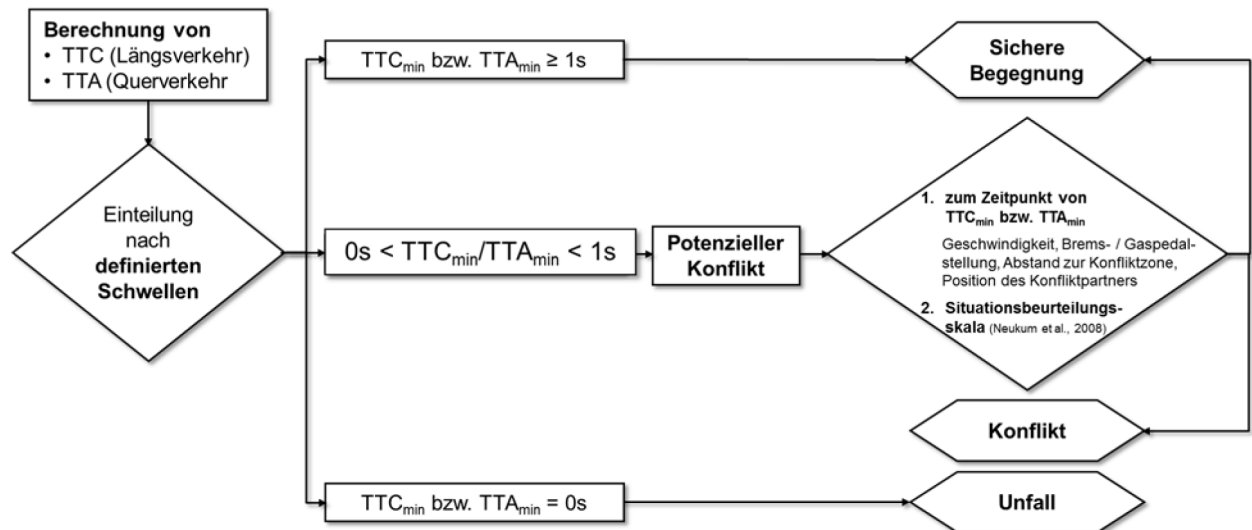


Abb.3: Klassifizierung nach Unfällen, Konflikten und sicheren Begegnungen.

Den Ausgangspunkt stellt die Berechnung des minimalen Zeitabstandes zwischen den Konfliktpartnern (Längsverkehr: Time-to-collisionmin (TTC = d/v_{rel}), Querverkehr: Time-to-arrivalmin (TTA = d/v)) dar. Ergibt sich als Minimum 0s, ereignet sich ein Unfall, bei Werten größer oder gleich 1s, eine sichere Begegnung. Werte zwischen 0s und 1s kennzeichnen potenzielle Konflikte (Hayward, 1972; van der Horst, 1990). Um unter den potentiellen Konflikten zwischen sicheren Begegnungen und tatsächlichen Konflikten zu unterscheiden, werden weitergehende Fahrdaten und das Urteil der Situationskritikalität herangezogen. Unfälle und Konflikte werden in der Ergebnisdarstellung zu ‚kritischen Begegnungen‘ zusammengefasst.

Hypothesen: In Anlehnung an Forschungsergebnisse zu Fahrerwarnungen wird erwartet, dass die Effektivität der **Fahrerunterstützung** von situationsspezifischen **Fahrererwartungen** abhängt: In Szenarien mit niedriger (vs. hoher) Erwartbarkeit sollte ein höherer Unterstützungsbedarf vorliegen (d.h. höherer Anteil kritischer Begegnungen), da die Fahrer nicht die Möglichkeit haben antizipativ auf die Konfliktsituationen zu reagieren. Weiterhin wird erwartet, dass eine **Verdeckung** der Konfliktpartner zu einem vermehrten Auftreten von Verkehrskonflikten führt.

Ergebnisse

Anteil kritischer Begegnungen: In nicht-assistierte Fahrten findet sich insgesamt eine Steigerung der Häufigkeit kritischer Begegnungen durch Verletzung der **Antizipationsmöglichkeit** von Konfliktsituationen (Erwartung niedrig: 163 von 403; hoch: 15 von 437 Begegnungen; $\chi^2_{df1}=172.00, p < .000$) und bei **Verdeckung der jeweiligen Konfliktpartner** (verdeckt: 103 von 423; sichtbar: 75 von 417¹ Begegnungen; $\chi^2_{df1}=5.09, p < .050$). Die grauen Balken in Abb.4 zeigen in abstei-

¹ Unterschiede in Randhäufigkeiten aufgr. fehlender Werte, da Prüfsituation nicht in geplanter Weise zustande kam.

Fahrerseitiger Unterstützungsbedarf in drohenden Verkehrskonfliktszenarien und Wirksamkeitsuntersuchung frühzeitiger Fahrerinformationen basierend auf kooperativer Umfelderkennung
 gender Häufigkeitsreihenfolge die zehn Szenarienvarianten, in denen es ohne Fahrerunterstützung zu kritischen Begegnungen kommt.

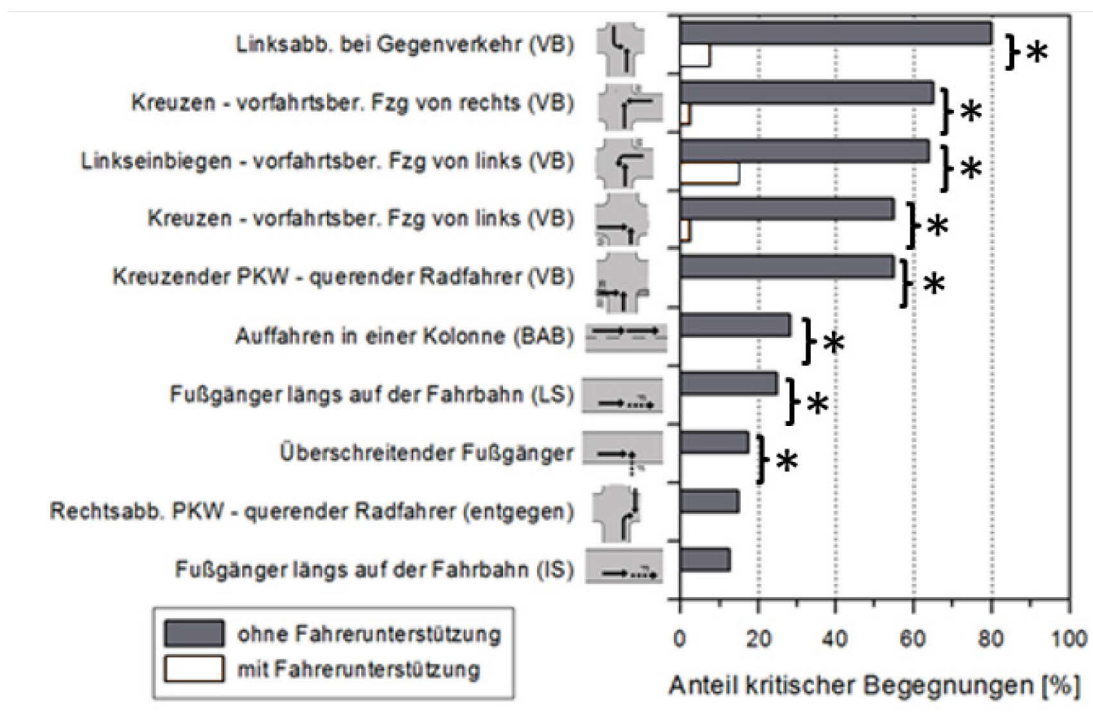


Abb.4: Anteil kritischer Begegnungen. Legende: VB = unterstützter Fahrer ist vorfahrtsberechtig; BAB = Bundesautobahn; LS = Landstraße; IS = Innenstadt; entgegen = Radfahrer nähert sich aus entgegengesetzter Fahrtrichtung; * = $p < .05$.

Die Effizienz einer durch kooperative Umfelderkennung ermöglichten Fahrerunterstützung zeigt sich in einer deutlichen Reduktion des Anteils kritischer Begegnungen in nahezu allen zehn Szenarienvarianten (mit Unterstützung: 11 von 350, ohne 143 von 353 Begegnungen: $\chi^2_{df1}=89.2, p < .000$).

Nachbefragungsergebnisse: Tab.4 zeigt die Nachbefragungsergebnisse. Die Ergebnisse spiegeln die angenommenen Erwartungseffekte wider: Situationen, in denen den Fahrern die Vorfahrt genommen wird (bspw. Variante „VB“ in Kreuzungsszenarien), oder in denen die Konfliktsituationen plötzlich und ohne Möglichkeit der Antizipation auftauchen (bspw. „Überschreiten“), erreichen die höchsten Einschätzungen der Unfallträchtigkeit. Betrachtet man die Nachbefragungsergebnisse zum subjektiven Nutzen der Fahrerunterstützung, so fällt auf, dass diese in allen Szenarien positiv bewertet wird: Die Mehrheit der Fahrer (> 70%) gibt an, dass das Fahren mit Fahrerunterstützung die Fahrsicherheit gesteigert hat.

Tab.4:Nachbefragung – Zustimmung zu den Nachbefragungsisems (in Prozent).

Szenario	Vorfahrt/ Fahrtrichtung	Vorkommen	Unfallträchtigkeit	Fahren sicherer
Auffahren in einer Kolonne		95	85	85
Überholen		95	50	75
Linksabbiegen bei Gegenverkehr	WP	85	47	75
	VB	15	58	95
Vorbeifahren bei Gegenverkehr		70	39	95
Kreuzen – vorfahrtsberechtigtes Fzg von links bzw. rechts	WP	95	32	70
	VB	35	61	85
Linkseinbiegen – vorfahrtsberechtigtes Fzg von links	WP	90	33	80
	VB	48	72	95
Überschreiten		70	72	95
Rechtsabbiegender PKW – querender FG		95	50	84
Linksabbiegender PKW – querender FG		90	35	85
FG in Längsrichtung auf der Fahrbahn		37	22	90
Kreuzender PKW – querender RF	WP	90	29	80
	VB	60	61	95
Rechtsabbiegen – querender RF	gleich	95	65	100
	entgegen	60	39	95

Diskussion

Erwartungsgemäß ergab sich in der vorliegenden Untersuchung ein fahrerseitiger Unterstützungsbedarf besonders in Situationen, die plötzlich auftreten und bei denen dem Fahrer die Vorfahrt durch andere Verkehrsteilnehmer genommen wird (beispielweise bei Kreuzungssituationen mit einem Anteil kritischer Begegnungen von mindestens 50%). In diesen Situationen werden Fahrererwartungen verletzt, wodurch eine Fahrerunterstützung notwendig wird. Die Untersuchung macht deutlich, dass die teilnehmenden Fahrer ihre Aufmerksamkeit auf Basis des jeweiligen situativen Kontexts auf mögliche Konfliktquellen ausrichteten und dementsprechend ihr Fahrverhalten anpassten. Dies entspricht den angeführten Literaturbefunden. Die Ergebnisse verdeutlichen das Potenzial kooperativer Umfelderkennung zur Entschärfung von unerwarteten und plötzlich auftauchenden Verkehrskonflikten. Insbesondere bei Kreuzungskonflikten konnte der hohe Anteil an Konfliktsituationen drastisch gesenkt werden. Weiterer Forschungsbedarf ergibt sich bezüglich der zu erwartenden Unsicherheit der frühzeitigen Fahrerinformationen.

Danksagung

Diese Arbeit entstand im Verbundprojekt Ko-PER, im Rahmen der vom Bundesministerium für Wirtschaft und Technologie geförderten Forschungsinitiative Ko-FAS (www.kofas.de) unter dem Förderkennzeichen 19S9022C.

Literatur

- Bliss, J. P. & Acton, S. A. (2003). Alarm mistrust in automobiles: how collision alarm reliability affects driving. *Applied Ergonomics*, 34(6), 499-509.
- Brown, T. L., Lee, J. D. & McGehee, D. V. (2000). Attention based model of driver performance in rear-end-collisions. *Transportation Research Record*, 1724, 14-21.
- Brown, T. L., Lee, J. D. & McGehee, D. V. (2001). Human performance models and rear-end collision avoidance algorithms. *Human Factors*, 43(3), 462-482.
- Campbell, J. L., Carney, C. & Kantowitz, B. H. (1997). Human factors design guidelines for advanced traveler information systems (ATIS) and commercial vehicle operations (CVO) (Technical Report FHWA-RD-98-057). Washington, D.C.: U.S. Department of Transportation, Federal Highway Administration.
- COMSIS Corporation. (1996). Preliminary human factors guidelines for crash avoidance warning devices (NHTSA Project No. DTNH22-91-07004). Silver Spring, MD: COMSIS.
- Dingus, T. A., Jahns, S. K., Horowitz, A. D. & Knipling, R. (1998). Human factors design issues for crash avoidance systems. In N. V. Barfield & A. Dingus (Eds.), *Human factors in intelligent transportation systems* (pp. 55-94). Mahwah, NJ: Erlbaum.
- GDV. (1998). *Unfalltypenkatalog: Leitfaden zur Bestimmung des Unfalltyps*. Köln: Institut für Straßenverkehr - Gesamtverband der deutschen Schadenversicherer e.V. .
- Green, P., Levison, W., Paelke, G. & Serafin, C. (1993). *Preliminary human factors guidelines for driver information systems* (Technical Report UMTRI-93-21). Ann Arbor, MI: The University of Michigan Transportation Research Institute.
- Hayward, J. C. (1972). *Near miss determination through use of a scale of danger*: The Pennsylvania State University, Pennsylvania.
- International Harmonized Research Activities (IHRA) working group on Intelligent Transport Systems (ITS). (2008). Statement of principles on the design of high-priority warning signals for in-vehicle intelligent transport systems – draft. Informal Document Number ITS-16-03.
- Koornstra, M. J. (1993). Safety relevance of vision research and theory. In A. G. Gale (Ed.), *Vision in Vehicles IV* (pp. 3-13). Amsterdam: Elsevier.
- Lenné, M. G. & Triggs, T. J. (2009). Warning drivers of approaching hazards: The importance of location cues and multi-sensory cues. In D. De Waard, J. Godthelp, F. L. Kooi & K. Brookhuis (Eds.), *Human Factors, Security and Safety* (pp. 203-211). Maastricht: Shaker Publishing.

- Muhrer, E. & Vollrath, M. (2010). Expectations while car following - the consequences for driving behaviour in a simulated driving task. *Accident Analysis and Prevention*, 42, 2158-2164.
- Naujoks, F., Grattenthaler, H. & Neukum, A. (2012). Zeitliche Gestaltung effektiver Fahrerinformationen zur Kollisionsvermeidung auf der Basis kooperativer Perzeption. In K. Dietmayer, K. Bengler, B. Färber, M. Maurer, C. Stiller & H. Winner (Eds.), 8. *Workshop Fahrerassistenzsysteme – FAS 2012* (pp. 107-117). Darmstadt: Uni-DAS.
- Neukum, A. (2011, 28.September). Wenn das Fahrzeug mehr sieht als der Fahrer – Konsequenzen für die Gestaltung der Fahrer-Fahrzeug Schnittstelle. Paper presented at the Ko-FAS Zwischenpräsentation, Schloss Aschaffenburg.
- Rhede, J. G., Wäller, C. & Oel, P. (2011). Der FAS Warnbaukasten – Strategie für die systematische Entwicklung und Ausgabe von HMI-Warnungen. In VDI Wissensforum (Ed.), 6. *VDI-Tagung - Der Fahrer im 21. Jahrhundert Fahrer, Fahrerunterstützung und Bedienbarkeit (VDI-Berichte 2134)* (pp. 63-75). Düsseldorf: VDI Verlag GmbH.
- Schmidt, G. & Krüger, H.-P. (2010). Violating the driver's threat anticipation and the effect of different modalities for Forward Collision Warnings (FCW), *HFES Europe Chapter Conference*. Berlin.
- Schweitzer, N., Apter, Y., Ben-David, G., Liebermann, D. G. & Parush, A. (1995). A field study on braking responses during driving, II. Minimum driver braking times. *Ergonomics*, 38, 1903 -1910.
- Sorkin, R. D. (1988). Why are people turning off our alarms? *Journal of the Acoustic Society of America*, 84(3), 1107-1108.
- Theeuwes, J. & Hagenzieker, P. (1993). Visual search of traffic scenes: On the effect of location expectations. In A. G. Gale (Ed.), *Vision in Vehicles IV* (pp. 149-158). Amsterdam: Elsevier.
- van der Horst, A. R. A. (1990). *A time-based analysis of road user behaviour in normal and critical encounters*. Dissertation, Technische Universiteit Delft, Delft, Niederlande.
- Van der Hulst, M., Meijman, T. F. & Rothengatter, J. A. (1999). Anticipation and the adaptive control of safety margins in driving. *Ergonomics*, 42, 336–345.
- Weidl, G. & Breuel, G. (2012). Overall Probabilistic Framework for Modeling and Analysis of Intersection Situations, *16th International Forum on Advanced Microsystems for Automotive Applications (AMAA 2012)*. Berlin.
- Winner, H. (2009). Frontalkollisionsschutzsysteme. In H. Winner, S. Hakuli & G. Wolf (Eds.), *Handbuch Fahrerassistenzsysteme* (pp. 522-542). Wiesbaden: Vieweg & Teubner.

Die dunkle Seite des Lichts - Diskomfort durch Flicker bei (LED-)Lichtern im Straßenverkehr in Bezug zu peripheren Flimmerverschmelzungsfrequenzen

Jens Mühlstedt, Patrick Roßner und Angelika C. Bullinger ¹

Schlüsselwörter: Flicker, Flimmern, Sehen, Flimmerverschmelzungsfrequenz, Diskomfort

Zusammenfassung

Leuchtdioden (light-emitting diode, LED) werden seit einigen Jahren mit hohen Wachstumsraten in allen Bereichen moderner Lichttechnik eingesetzt. Auch im Straßenverkehr, insbesondere bei Lichtzeichenanlagen wie z. B. Ampeln und in Scheinwerfern, kommen LEDs zum Einsatz. Dies hat verschiedene Vorteile, wie hoher Wirkungsgrad, längere Lebensdauer, kurze Ansprechzeiten und günstiges Defektverhalten. Speziell im Straßenverkehr steht diesen jedoch ein wesentlicher Nachteil entgegen: durch eine gepulste Ansteuerung der LEDs entsteht ein optischer Flicker: ein hochfrequentes Pulsieren der Lichter, das Diskomfort auslösen kann.

Der Beitrag steht unter der Forschungsfrage des Diskomforts durch Flicker. Analysiert wird der subjektive Diskomfort im Bezug zu objektiven Flimmerverschmelzungsfrequenzen sowie Unterschiede von zentralen und peripheren Flimmerverschmelzungsfrequenzen. Die erste von zwei durchgeführten Studien hat zum Ergebnis, dass etwa ein Drittel der Probanden den Flicker wahrgenommen haben und davon etwa die Hälfte diesen als (sehr) unangenehm empfinden. Die periphere Flimmerverschmelzungsfrequenz differiert nach der zweiten Studie von der zentralen Flimmerverschmelzungsfrequenz und hat im Bereich zwischen 20° und 30° um den zentralen Sehpunkt um ca. 10 Hz höhere Frequenzen. Die zentrale Flimmerverschmelzungsfrequenz beträgt $47 \text{ Hz} \pm 5 \text{ Hz}$, die peripheren betragen um 25° etwa 54-56 Hz.

Die Ergebnisse zeigen sehr deutlich, dass der Flicker seit der Nutzung von LEDs in Verbindung mit gepulster Ansteuerung ein relevantes Phänomen im Straßenverkehr ist. Technische sowie ergonomische Fragestellungen sind offen und bedürfen dringend einer Untersuchung. Insbesondere die Frage, wie es zu dem Diskomfort bei Frequenzen von 100-200 Hz verursacht durch Ampeln und Fahrzeugleuchten kommt, sollte beantwortet und in Vorgaben für die Leuchtenhersteller überführt werden.

¹ Technische Universität Chemnitz, Professur Arbeitswissenschaft und Innovationsmanagement

Die dunkle Seite des Lichts - LEDs und Flicker

Die technische Basis künstlicher Beleuchtung befindet sich in einem Umbruch. Seit der Umsetzung der Energieeffizienzklassen in der EU und damit einhergehendem Herstellungs- und Vertriebsverbot für Glühlampen (Richtlinie 2009/125/EG, 2009) findet neben dem Ersatz herkömmlicher Leuchtmittel durch sogenannte Retrofits ein Einsatz von LEDs in verschiedensten Anwendungen statt (Fördergemeinschaft Gutes Licht, 2010). Von der Hintergrundbeleuchtung von Monitoren und Fernseher über die Nutzung in Schreibtisch- und Raumbeleuchtungen bis hin zum Einsatz in automotiven Anwendungen ergeben sich mehrere Vorteile, die durch die neue Technik genutzt werden können (Tab. 1). Diesen stehen nur wenige Nachteile gegenüber, denen aber besondere psycho-physiologische Relevanz innewohnt (Tab. 1). Selbst ein vermeintlicher Vorteil - die geringe Einschaltdauer - kann in Verbindung mit einer Pulsung zur Verstärkung negativer Effekte führen.

Tab. 1: mögliche Vor- und Nachteile von LED-Leuchtmitteln gegenüber konventionellen Leuchtmitteln (Glühlampe, Halogenlampe, Gasentladungslampe)

Vorteile von LEDs	Nachteile von LEDs
lange Lebensdauer hohe Energieeffizienz beliebige (Licht-)Farben vibrationsunempfindlich geringe Wärmeentwicklung frei von IR- und UV-Strahlung Dimmbarkeit geringe Einschaltdauer (Fördergemeinschaft Gutes Licht, 2010)	notwendige Elektronik (Hanus, 2004) z. T. Höhere Anschaffungskosten (Fördergemeinschaft Gutes Licht, 2010) punktuelle Wärmeentwicklung (Krückeberg, 2007) punktuell hohe Leuchtdichten (BAuA, 2013) Flicker

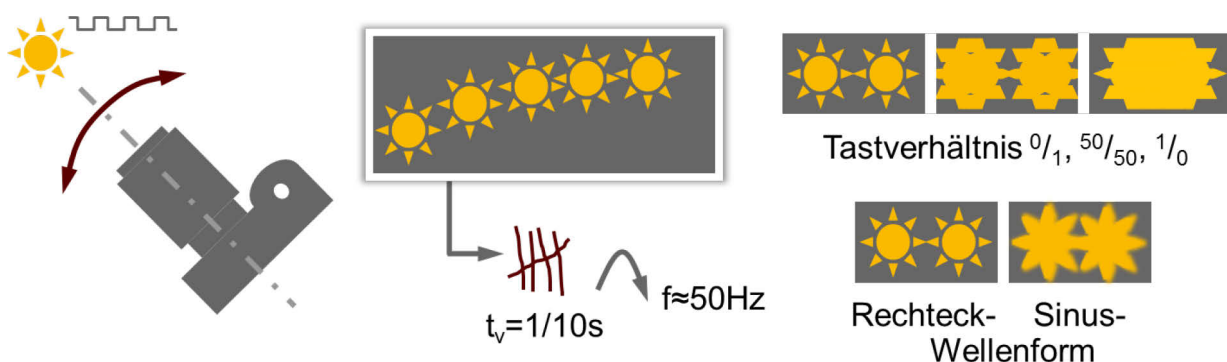
In verschiedenen Anwendungen werden LEDs nicht mit einer konstanten Gleichspannung versorgt, sondern mit unterschiedlichen Frequenzen sowie aus bestimmten Gründen gepulst angesteuert (Tab. 2, Abb. 1). Ein Grund ist die Gleichrichtung von Wechselströmen zum Betreiben der Dioden. Weitere Ursachen sind die Erzeugung eines Tastverhältnisses zur Dimmung, Wärme- oder Leistungsregulation.

Tab. 2: Beispiele für die Pulsung von LED-Licht (eigene Messungen)

Anwendung	exemplarische Frequenzen
Ampeln / Lichtzeichenanlagen	50 Hz, 100Hz
KFZ-Rücklichter	100Hz, 200Hz, 600Hz
KFZ-Tagfahrleuchten	200 Hz
KFZ-Frontscheinwerfer	200 Hz, 600 Hz
KFZ-Innenraumbeleuchtung	75 Hz, 200 Hz, 500 Hz
Wechselverkehrszeichen	70 Hz, 300 Hz, ∞ Hz
LED-Info-Anzeigedisplays	50 Hz, 100 Hz, 400 Hz

**Abb. 1: Beispiele für Flicker in Rückleuchten (links), Wechselverkehrszeichen (Mitte) und Anzeigen, hier eines Fahrstuhls (rechts)**

Zur Ermittlung von Flicker sind Flickermeter erhältlich, die Frequenz, Tastverhältnis usw. ermitteln (z. B. Klein, 2013; Arri, 2013). Eine einfache Methode zur überschlägigen Ermittlung der Flickerfrequenzen von Leuchtquellen sowie für schwer erreichbare Objekte ist die Nutzung digitaler Kameras. Bei bekannter Verschlusszeit und einer Relativbewegung der Kamera zum Objekt (Schwenken) erscheint die Leuchtquelle mehrfach auf dem Bild, man kann von einer Perlschnur sprechen (Abb. 2). Bei mehrfacher Durchführung ist die Flickerfrequenz hinreichend genau ermittelbar und eine grobe Einschätzung von Tastverhältnis und Wellenform möglich.

**Abb. 2: Ermittlung von Flickerfrequenzen mit Digitalkameras (Beispiel: bei einer Verschlusszeit von 1/10 s und fünf auf dem Bild sichtbaren Objekten ergibt sich eine Flickerfrequenz von etwa 50 Hz)**

Die Flimmerverschmelzungsfrequenz des menschlichen Auges - Ich sehe was, was Du nicht siehst

Der menschliche Sehapparat ist das wichtigste Wahrnehmungsorgan (Mühlstedt, Glöckner, Spanner-Ulmer, 2011). Ein Großteil der Informationen wird visuell aufgenommen. Hierbei erfordern technische Innovationen der Interaktion immer wieder die Überprüfung notwendiger Anpassung an den menschlichen Sehapparat, der bestimmte Eigenschaften und Fähigkeiten besitzt (Roßner, Hartwich, Bullinger, 2012). Eine der wichtigen Eigenschaften des Auges, neben Akkomodations-, Auflösungs- und Adaptationsvermögen sowie Farbwahrnehmung, ist die Flimmerverschmelzungsfrequenz (engl. flicker-fusion threshold). Diese bezeichnet "die Frequenz, bei der eine Folge von Lichtblitzen als ein kontinuierliches Licht wahrgenommen wird" (Schlick, Luczak, Bruder, 2010). Es wurden verschiedene Parameter beschrieben, die Einfluss auf diese Frequenz haben (siehe insbesondere Landis, 1954; Kleberger, 1960):

- Leuchtdichte (Jaklitsch, 2004) und Signal-Noise-Ratio / SNR (Kleberger, 1960), Wellenlänge und Tastverhältnis (Groneberg, 2010), Form und Größe des Reizes (Kleberger, 1960),
- Adaptationszustand der Netzhaut, Vigilanz d. Probanden (Schmidtke, 1965),
- Alter der Probanden (Lachenmayer et al., 1994; Groneberg, 2010),
- Umgebungsfaktoren wie Hintergrundgeräusch (Amelang et al., 2006).

Einige in bisherigen Untersuchungen ermittelte Flimmerverschmelzungsfrequenzen sind in Tab. 3 angegeben. Die großen Unterschiede und Bandbreiten entstehen durch veränderte Parameter, die je nach Versuchsbedingung verschiedene Frequenzen ergeben.

Tab. 3: Flimmerverschmelzungsfrequenzen aus der Literatur

Frequenz	Autor(en)	Angegebene Parameter
0-55Hz, bis 85 Hz	Landis, 1954	0-55 Hz zentral, bis 75 Hz peripher, bis 85 Hz bei best. Lichtfarben
0-97 Hz	Hentschel, 2002	in Abhängigkeit von z. B. Umfeldleuchtdichte
10/25-80 Hz, meist 50-60 Hz	Tauer, 2010	10-25 Hz bei geringen Leuchtdichten, 80 Hz bei großer Helligkeit, meist 50-60 Hz
20-85 Hz	Ehrenstraßer, 2008	
82 Hz	Hylkema, 1942	30° großes Feld, 500 cd/m ²
bis 85 Hz	Gralla, 2007	

Ein bislang wenig untersuchter Parameter ist die Lage des Reizes auf der Netzhaut, die periphere Flimmerverschmelzungsfrequenz. Hierzu wurden Angaben zum

ungefähren Verlauf gemacht - dieser sei im Bereich von 30-40° um die Mitte des Gesichtsfeldes am größten und sowohl zur Mitte als auch nach außen geringer (Roehrig, 1959; Gralla, 2007). Ein weiterer wenig untersuchter Parameter ist die Streuung der Frequenzen in Populationen. Die Untersuchungen der Professur Arbeitswissenschaft und Innovationsmanagement setzen an diesen Forschungslücken - periphere Flimmerverschmelzungsfrequenz und Streuung dieser in Populationen - an und untersuchen diese. Doch zuerst soll untersucht werden, wie hoch die Relevanz des Flickers im Straßenverkehr ist.

Studie zu Diskomfort durch Flicker

Zur Beantwortung der Forschungsfrage, ob Flicker von Lichtquellen im Straßenverkehr erkannt wird und einen Diskomfort hervorruft, wird eine quantitative Studie zur Wahrnehmung sowie subjektiver Bewertung des Flickers bei Lichtzeitanlagen und Fahrzeuglichtern durchgeführt. Für die Studie wurden insgesamt 1.792 Personen mittels eines standardisierten Fragebogens befragt, ein Teil davon online und ein Teil offline auf Parkplätzen. Die Probanden waren im Durchschnitt $34,7 \pm 12,0$ Jahre alt, zu 47% weiblich und 55% nutzen eine Sehhilfe.

Die Auswertung zeigt, dass 40% der Probanden ein Flimmern bei Ampeln wahrgenommen haben, 22% bei Fahrzeugrückleuchten und 29% bei Tagfahrleuchten (Abb. 3). Die statistische Auswertung zeigt, dass weder Alter noch Geschlecht, Sehhilfe oder Teilnahme am Straßenverkehr einen signifikanten Einfluss auf diese Häufigkeiten haben.

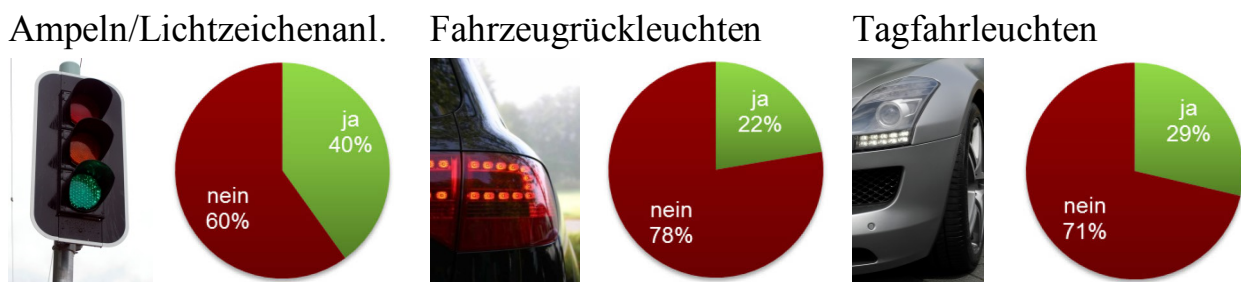


Abb. 3: Wahrnehmung eines Flimmerns/Flackerns bei den erfragten LED-Lichtquellen
(Bildquellen: Wikimedia Commons, 2013)

Das Flimmern wird von jeweils über der Hälfte der Befragten als unangenehm oder sehr unangenehm beurteilt (Ampeln 55% / Rückl. 63% / Tagfahrl. 58%) und wird nicht zu einer bestimmten Tageszeit verstärkt wahrgenommen (ähnliche Anteile fallen auf tagsüber / Dämmerung / nachts und "weiß nicht").

Die Anzahl der Menschen, die das Flimmern in den untersuchten Anwendungen bereits wahrgenommen haben, ist überraschend hoch. Gingen Prognosen von einstelligen Prozentzahlen aus, legen Werte von 20-40% eine hohe Relevanz des Themas dar. Die Beurteilung des Flimmerns hingegen ist wie erwartet sehr negativ, und das interessanterweise unabhängig von der Tageszeit. Schlussfolgernd

stellt sich die Forschungsfrage, inwiefern diese subjektive Alltagswahrnehmung mit konkreten Laboruntersuchungen zur Flimmerverschmelzungsfrequenz nachvollzogen werden kann. Folgend wird dazu eine zweite Studie vorgestellt.

Studie zu peripheren Flimmerverschmelzungsfrequenzen

Um periphere Flimmerverschmelzungsfrequenzen des Menschen zu ermitteln, wurde eine Laborstudie durchgeführt. Ein Versuchsträger, der gepulstes Licht variabler Frequenz (10-300 Hz) abgibt, wird in verschiedenen Punkten des Gesichtsfeldes abgebildet, indem nacheinander verschiedene Punkte im Raum fokussiert werden. Der Versuchsleiter ermittelt die jeweilige Flimmerverschmelzungsfrequenz des Probanden durch absteigendes und aufsteigendes Ändern der Pulsfrequenz. Als Versuchsträger dient ein LED-Fahrzeug-Rücklicht (Abb. 4). In diesem sind 12 rote LEDs zweireihig angeordnet, sodass sich mit einem Abstand zu den Augen der Probanden von 4,1 m und einer Leuchtfläche von 110x30 mm² ein Sehinkel der Lichtquelle von 1,5°x0,4° ergibt. Die Lichtquelle besitzt eine Leuchtdichte von $L=2.000 \text{ cd/m}^2$ und die Umgebung im Mittel $L=3 \text{ cd/m}^2$, was einem Leuchtdichteverhältnis von etwa 1:700 entspricht. Ausgehend von der Lichtquelle sind 50 Messpunkte in 8 Strahlen bis zu den Grenzen des Gesichtsfeldes angeordnet (Abb. 4/5). An dem Versuch nahmen $n=28$ Probanden teil. Das Alter der Probanden beträgt $25,3 \pm 4,7$ Jahre, 30% sind weiblich und 41% nutzen eine Sehhilfe.



Abb. 4: Versuchsträger (links, Leuchtelemente sind die 12 LEDs auf der linken Seite, Frequenz- und Tastverhältnis über Bedienfeld links) und Versuchsaufbau (rechts - Versuchsträger mit Versuchsleiter links oben, Proband rechts unten)

Die ermittelten Flimmerverschmelzungsfrequenzen liegen im Probandenmittel zwischen 38 und 56 Hz. Es zeigt sich eine Verteilung, wie sie auch in anderen Untersuchungen beschrieben wurde: im zentralen Gesichtsfeld ist diese eher niedrig mit 47 Hz, steigt peripher bis etwa 20°-30° im Gesichtsfeld auf 54-56 Hz an, um nach außen bis an die Ränder des Gesichtsfeldes unter 50 Hz abzufallen (Abb. 5).

Die Frequenzen streuen intraindividuell um etwa $s=5,6 \text{ Hz}$, wobei die Daten normalverteilt sind und das 5./95. Perzentil je 9 Hz vom Mittelwert und das 1./99.

Perzentil etwa 13 Hz vom Mittelwert entfernt sind. Das absolute Maximum der Studie liegt bei einer Frequenz von 87,5 Hz.

Die peripheren Flimmerverschmelzungsfrequenzen dienen nicht als Erklärung für den Diskomfort aus der ersten Studie. Hierzu sind weitere Forschungsarbeiten notwendig, denn insbesondere die maximalen Flimmerverschmelzungsfrequenzen um 100 Hz stehen Perlschnurverschmelzungsfrequenzen von 300-900 Hz gegenüber (Strauß, 2007). Dies sind Grenzfrequenzen, bei denen ein Netzhautabbild einer Lichtquelle als unterbrochene Linie wahrgenommen wird, das sich relativ auf dieser bewegt. Es existiert eine Forschungslücke bezüglich "dynamischer", eine Relativbewegung der Lichtquelle im Gesichtsfeld aufweisender Flimmerverschmelzungsfrequenzen. Was im Detail zwischen den Flimmer- und den Perlschnurverschmelzungsfrequenzen für Effekte auftreten, muss weiter wissenschaftlich untersucht werden.

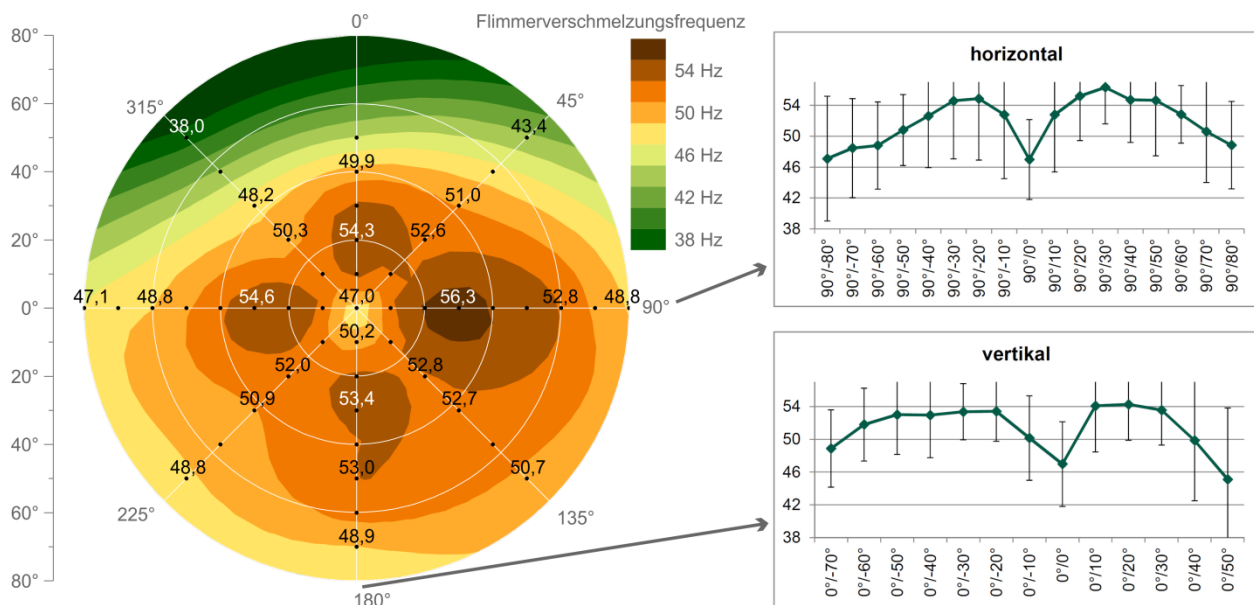


Abb. 5: periphere Flimmerverschmelzungsfrequenzen im Flächendiagramm (links) und in zwei Schnitten (rechts)

Zusammenfassung und Ausblick

Moderne Lichttechnik bringt durch den Einsatz von LEDs nicht nur Vorteile mit sich, sondern auch einige Probleme, wie den Flicker. Dieser kann unter bestimmten Bedingungen Diskomfort durch unvollständige Flimmerverschmelzung erzeugen. Eine dieser Bedingungen ist das Auftreten des Reizes im peripheren Gesichtsfeld. Dieser erklärt den Effekt aber noch nicht vollständig. Eine weitere Bedingung, die sich in Beobachtungen zeigt, scheint eine Bewegung des Reizes im Gesichtsfeld zu sein, verursacht entweder durch Bewegungen des Objekts oder der Augen des Betrachters. Hier sollten Folgeforschungen anknüpfen.

Weiterhin ist die Frage offen, ob durch ein wahrgenommenes Flimmern ein Diskomfort zustande kommt oder sogar die Sicherheit im Straßenverkehr gefährdet ist. Strategien zur Vermeidung des Flimmerns neben einer nicht immer sinnvollen

Frequenzerhöhung (mögliche Probleme zur elektromagnetischen Verträglichkeit / EMV) sollten ebenfalls gesucht werden. Darüber hinaus existieren weitere Detailfragen, die in diesem Zusammenhang erforscht werden sollten, etwa nichtvisuelle Auswirkungen von Flicker, der Wellenform (Rechteck, Sägezahn, Sinus etc.) sowie mögliche Effekte beim Auftreten mehrerer Flicker-Lichtquellen.

Der Beitrag entstand z. T. auf Basis der Projektarbeit von Tim Heidrich (2013).

Literatur

- Amelang, Manfred; Bartussek, Dieter; Stemmler, Gerhard; Hagemann, Dirk. (2006). *Differentielle Psychologie und Persönlichkeitsforschung*. Kohlhammer.
- Arri. (2013). Light Analyzer. Zugriff am 25.07.2013: http://www.arri.com/lighting/americas/arri_accessories/light_analyzer.html.
- BAuA. (2013). Photobiologische Sicherheit von Licht emittierenden Dioden (LED). Forschung Projekt F 2115. Dortmund/Berlin/Dresden: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. ISBN 978-3-88261-726-9.
- Ehrenstraßer, Michael. (2008). Sensoreinsatz in der telepräsenten Mikromontage. Herbert Utz Verlag.
- Fördergemeinschaft Gutes Licht. (2010). *Licht.wissen 17. LED: das Licht der Zukunft*. ISBN-Nr. Druckausgabe 978-3-926193-55-1.
- Gralla, V. (2007). Peripheres Sehen im Sport - Möglichkeiten und Grenzen dargestellt am Beispiel der synchroptischen Wahrnehmung. Bochum: Ruhr-Universität Bochum.
- Groneberg, T. (2010). Die Flimmerverschmelzungsfrequenz bei altersabhängiger Makuladegeneration. München: Technische Universität München.
- Hanus, B. (2004). *Experimente mit superhellen Leuchtdioden*. Poing: Franzis Verlag.
- Heidrich, Tim. (2013). Untersuchung des Diskomforts von gepulsten LED-Leuchten. Projektarbeit. Technische Universität Chemnitz.
- Hentschel, Hans-Jürgen. (2002): *Licht und Beleuchtung*. 5., neubearb. und erw. Aufl. Hüthig, Heidelberg.
- Hylkema, B. S. (1942). Fusion Frequency with Intermittent Light under Various Circumstances. *Acta ophthal. (Kbh.)*, 20, 159-180.
- Jaklitsch, Walter (2004). Handbuch der Laufbildfotografie. Berlin, Heidelberg: Springer.
- Kleberger, Eberhard. (1960). Zur Pathologie der Flimmerverschmelzungsfrequenz. *Documenta Ophthalmologica*, 1961, 5(1), S. 226-320.
- Klein. (2013). Klein K10-A Colorimeter. Zugriff am 25.07.2013: <http://www.kleininstruments.com/colorimeters/k10-a/>.

- Krückeberg, J. (2007). *Hochleistungs-LEDs in der Praxis*. Poing: Franzis Verlag GmbH.
- Lachenmayer, B., Kojetinsky, S., Ostermaier, N., Angstwurm, K., Vivell, P., & Schaumberg, M. (Mai 1994). The Different Effects of Aging on Normal Sensitivity in Flicker and Light-Sense Perimetry. *Investigative Ophthalmology & Visual Science*, 35(6), S. 2741-2748.
- Landis, Carney. (1954). Determinants of the Critical Flicker-Fusion Threshold. *Physiol. Rev.*, 34(2), S. 259-286.
- Mühlstedt, Jens; Glöckner, Stev; Spanner-Ulmer, Birgit. (2011). Licht und Farbe am Wissensarbeitsplatz – ergonomische Anforderungen der Nutzer. In: Gesellschaft für Arbeitswissenschaft e. V. (Hrsg.). *Mensch, Technik, Organisation - Vernetzung im Produktentstehungs- und -herstellungsprozess*, Bericht zum 57. Arbeitswissenschaftlichen Kongress, S. 331 - 334. Dortmund: GfA-Press. ISBN 978-3-936804-10-2.
- Richtlinie 2009/125/EG. (2005). Richtlinie 2009/125/EG des Europäischen Parlaments und des Rates vom 21. Oktober 2009 zur Schaffung eines Rahmens für die Festlegung von Anforderungen an die umweltgerechte Gestaltung energieverbrauchsrelevanter Produkte.
- Roehrig, W. (1959). The Influence of the Portion of the Retina stimulated on the Critical Flicker-Fusion Threshold. *The Journal of Psychology*, S. 57-63.
- Roßner, P.; Hartwich, F.; Bullinger, A. C. (2012). ViFa 65plus – Visuelle Fahrerassistenzsysteme zur Kompensation altersbedingter Veränderungen. Intelligent vernetzte Arbeits- und Fabrikssysteme, *Tagungsband VPP 2012 - vernetzt planen und produzieren & Symposium Wissenschaft und Praxis*. 8. November 2012 - 9. November 2012, Chemnitz, S. 405-410.
- Schlick, Christopher; Bruder, Ralph; Luczak, Holger. (2010). *Arbeitswissenschaft*. Heidelberg: Springer. ISBN 987-3-540-78332-9.
- Schmidtke, H. (1965). *Die Ermüdung*. Bern: Hans Huber.
- Strauß, Steffen. (2007). Theoretische und experimentelle Untersuchungen zum Einsatz gepulster Halbleiterlichtquellen in der Kraftfahrzeugbeleuchtung. Dissertation. Heinz Nixdorf Institut, Universität Paderborn.
- Tauer, Holger. (2010). *Stereo-3D*. Fachverlag Schiele & Schoen.
- Wikimedia Commons. (2013). File. Zugriff am 25.07.2013: http://upload.wikimedia.org/wikipedia/commons/4/47/LED_traffic_light.jpg, http://upload.wikimedia.org/wikipedia/commons/b/b0/A6_avant_c6_led.jpg http://upload.wikimedia.org/wikipedia/commons/thumb/1/1c/Mercedes-Benz_SLS_AMG_%28C_197%29_%E2%80%93_Scheinwerfer%2C_10._August_2011%2C_D%C3%BCsseldorf.jpg.

Visuelle Informationsaufnahme im virtuellen 3D-Raum

Elisabeth Dittrich¹, Anne Ostheimer² und Axel Vick³

Schlüsselwörter: virtueller dreidimensionaler Raum, 3D, Cave Automatic Virtual Environment, CAVE, Mensch-Maschine-Interaktion

Zusammenfassung

Neue Konzepte zur Visualisierung von Informationen sind z. B. Head-Up-Displays. Diese werden auch schon in den frühen Phasen der Cockpitentwicklung getestet. Dieses kann unter Anderem in virtuellen Umgebungen mit interaktiven 3D-Modellen geschehen. Daher stellt sich die Frage, inwieweit die Ergebnisse in die Realität übertragbar sind, da sich die menschliche Wahrnehmung im virtuellen Raum insbesondere durch den Konvergenz-Akkommodations-Konflikt (Hoffman et al., 2008), von jener im realen Raum unterscheidet. In einer Studie im realen Raum von Lichtenstein (2009) wurde der Einfluss des Darstellungsortes spezifischer Information auf die menschliche Aufnahme und -verarbeitung untersucht und eine aufgabenspezifische Verarbeitung in verschiedenen Tiefenebenen nachgewiesen: eine einfache Objektdetektionsaufgabe wurde auf der vorderen und eine numerische Rechenaufgabe auf der hinteren Ebene schneller verarbeitet. In Erweiterung dazu wird in dieser Studie die aufgabenspezifische Verarbeitung im virtuellen Raum überprüft. Die verwendete virtuelle Umgebung ist die CAVE vom Fraunhofer IPK. Im Fernbereich konnten Ergebnisse aus dem realen Versuch repliziert werden. Im Nahbereich jedoch nicht, was auf einen Einfluss des Konvergenz-Akkommodations-Konflikts schließen lässt.

Einleitung

Beim Einsatz von z. B. Head-Up-Displays in Fahrzeugen muss der Nutzer Informationen in verschiedenen Tiefenebenen wahrnehmen und verarbeiten, z. B. Geschwindigkeitsanzeige im Cockpit, Verkehrsinformationen auf dem Head-Up-Display auf der Frontscheibe und Umgebungsinformationen auf der vorausgehenden Fahrbahn. Das auf diesen verschiedenen Tiefenebenen die Informationen unterschiedlich schnell verarbeitet werden wurde von Previc (1998) thematisiert und u. A. von Lichtenstein (2009) auf zwei Tiefenebenen aufgezeigt.

Einer der beeinflussenden Parameter dafür ist Wahrnehmung der räumlichen Tiefe und diese wird durch verschiedene Tiefenhinweisreize ermittelt. Z. B. die okulomotorischen Hinweisreize Konvergenz (Anpassung der Augenstellung) und Akkommodation (Anpassung der Brechkraft durch Veränderung der Linsenform).

¹ Graduiertenkolleg prometei, TU Berlin FG Industrielle Informationstechnik, Fraunhofer IPK Geschäftsfeld Virtuelle Produktentstehung

² TU Berlin, FG Mensch-Maschine-Systeme

³ TU Berlin, FG Industrielle Automatisierungstechnik

Beim realen 3D-Sehen, sind diese gekoppelt, d.h. die Augen akkommodieren und konvergieren auf die gleiche Distanz. Im virtuellen Raum hingegen verhält sich das anders, denn diese Kopplung ist nicht vorhanden. Die Augen akkommodieren stets auf die Bildschirmoberfläche, auf der das Objekt tatsächlich physisch projiziert wird. Die Konvergenz ist allerdings auf den virtuellen Darstellungsort, d.h. auf die durch die 3D-Brille implizierte virtuelle Projektionsebene, bezogen. Dieser Konflikt zweier unterschiedlicher Tiefen-Hinweisreize kann zu einem Leistungseinbruch, visueller Ermüdung und Unbehagen führen (Hoffmann et al., 2008).

Daher ergibt sich die folgende Fragestellung: Tritt die aufgabenspezifische Verarbeitung in verschiedenen virtuellen Tiefenebenen ebenso wie im realen Raum auf? Denn durch die unterschiedlichen Tiefeninformationen der Konvergenz und Akkommodation könnte dieser Effekt der Tiefenwahrnehmung und der damit einhergehenden Informationsverarbeitung gestört werden. Um diese Frage zu beantworten wurde der Versuch von Lichtenstein (2009) in eine virtuelle Umgebung übertragen.

Methode

Es nahmen 22 Versuchspersonen (13 männlich, 9 weiblich) mit einem Durchschnittsalter von 28 Jahren teil. Davon hatten 73% keine Erfahrungen mit virtuellen 3D Umgebungen und 70% aller Versuchspersonen nahmen den Unterschied zwischen den Tiefenebenen wahr.

Bei der Studie kamen zwei Aufgabentypen zum Einsatz: eine einfache Objektdetektionsaufgabe, bei dieser musste auf das Erscheinen weißer Smileys, welche randomisiert alle eine bis neun Sekunden an einer beliebigen Stelle der jeweiligen, schwarz hinterlegten Projektionsfläche eingeblendet wurden, verbal (d.h. mit einem "Ja") reagiert werden. Die Reaktionszeiten – d.h. die Zeit zwischen Erscheinen des Smileys und verbaler Response der Versuchsperson – wurden hierbei durchgehend aufgezeichnet.

Und eine numerische Intelligenzaufgabe, welche darin bestand, einstellige Zahlen, welche jeweils nacheinander im Zentrum der Projektionsflächen visuell dargestellt wurden, zu summieren. Zu beachten war, dass der jeweils letzte Summand der vorherigen Rechnung den ersten Summanden der folgenden Rechnung bildete. Der erste Summand wurde eine bis neun Sekunden nach Programmstart eingeblendet, der zweite Summand wiederum eine bis neun Sekunden nach dem ersten. Die Summanden wurden jeweils drei Sekunden lang angezeigt. Auf den letzten Summanden folgte umgehend ein Ergebnisvorschlag mit Fragezeichen (z.B. „13?“), wobei dieser durch das Betätigen der linken bzw. rechten Taste einer Maus bestätigt bzw. negiert werden musste. Erst mit der Eingabe des Probanden über die Maus verschwand der Ergebnisvorschlag und die Reaktionszeit (Zeit zwischen Erscheinen des Ergebnisvorschlags und der Tasteneingabe) sowie die Richtigkeit der Antwort wurden gespeichert. Nach der Tasteneingabe wurde in einem zeitlichen

Abstand von einer bis neun Sekunden der nächste Summand angezeigt, welcher mit dem letzten Summanden der vorherigen Rechnung summiert werden sollte.

Die vordere Projektionsebene wurde in einem Abstand von einer Armlänge (ca. 65 cm) von der Versuchsperson (VP) entfernt dargestellt. Die hintere Projektionsebene war nochmals vier Meter von der vorderen Ebene (465 cm von der VP) entfernt. Die Versuchspersonen wurden so in der CAVE platziert, dass das Zentrum der Projektionsebenen jeweils auf Augenhöhe lag und ca. zwei Meter von der Bildschirmenebene entfernt saßen, d.h. die Augen akkommodierten stets auf diese feste Distanz (siehe Abbildung 1).

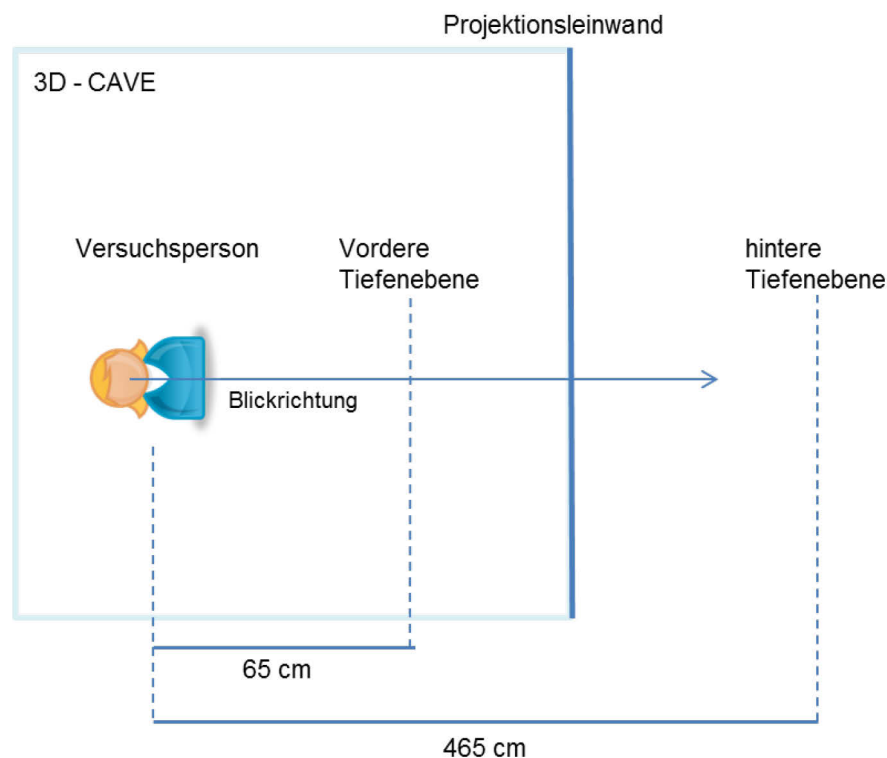


Abbildung 10: Versuchsaufbau in der CAVE

Die Studie setzte sich insgesamt aus acht Durchgängen zusammen - vier Baseline Trials (BTs) und vier Experimental Trials (ETs). In den Baseline Trials, welche jeweils zwei Minuten lang waren, wurden die beiden Aufgabentypen einzeln bearbeitet. Zunächst wurde die einfache Objektdetektionsaufgabe einmal auf der vorderen (BT1) und einmal auf der hinteren Projektionsebene (BT2) bearbeitet, anschließend folgte Gleiches mit der numerischen Intelligenzaufgabe (BT3/BT4). Ziel der Baseline Trials war die Gewöhnung der Versuchspersonen an den Versuchsaufbau und die beiden Aufgabentypen.

Während der Experimental Trials (ETs) von jeweils 3,5 Minuten Länge wurden beide Aufgabentypen parallel bearbeitet. In zwei von vier ETs wurden die Aufgabentypen auf unterschiedlichen Ebenen bearbeitet. In ET1 wurden die einfache Objektdetektionsaufgabe auf der vorderen und die numerische Intelligenzaufgabe auf der hinteren Ebene dargestellt, bei ET2 wurden die Ebenen getauscht, d.h. die nu-

merische Intelligenzaufgabe wurde vorne und die einfache Objektdetektionsaufgabe hinten bearbeitet. Zudem wurden die Aufgaben in zwei ETs auf derselben Ebene eingeblendet – entweder beides auf der vorderen (ET3) oder beides auf der hinteren Ebene (ET4). Die Reihenfolge der Experimental Trials wurde im Gegensatz zu den Baseline Trials randomisiert, um Reihenfolgeeffekte auszuschließen. Es handelte sich bei der gesamten Studie um ein Innersubjekt-Design, d.h. alle Versuchspersonen durchliefen alle Baseline und Experimental Trials. Die Darstellungsebenen der beiden Aufgabentypen waren während eines Trials konstant.

Technische Umsetzung

Die Studie wurde in der Cave Automatic Virtual Environment (CAVE) des Fraunhofer Instituts für Produktionsanlagen und Konstruktionstechnik (IPK) durchgeführt (vgl. Dittrich und Israel, 2012). Die CAVE bietet den maximalen Grad an Immersion bei möglichst ähnlichen Bedingungen wie in der oben genannten realen Studie von Lichtenstein (2009).

Zur dreidimensionalen Platzierung der Smileys und Summanden wurde die vom Institut entwickelte Erweiterung der CAVE um das TUI-Framework für die Interaktion mit Objekten (Israel et al., 2011) verwendet. Darauf aufbauend wurde ein Framework für frei programmierbare Reaktionszeitexperimente implementiert, das es ermöglicht, beliebige Objekte im Raum darzustellen und über unterschiedliche Modalitäten (z.B. Sprache, Maus -und Tastatureingaben) zu manipulieren.

Die zusätzliche Ausstattung zur Durchführung des Experimentes bestand aus einem Mikrofon mit eigenem Signalverarbeitungs-PC zur Verarbeitung der Audiokommandos, sowie einem Experimental-PC für das Hauptprogramm und der Mauseingabe. Die Audiokommandos zur Quittierung der Detektionsaufgabe der Smileys wurde über eine Netzwerkverbindung mit dem Hauptprogramm synchronisiert und dort über eine hochpräzise Reaktionszeitmessung ausgewertet. Die Verzögerung durch Signalverarbeitung und Netzwerkdatenübertragung sind stets konstant und können daher bei der qualitativen Auswertung der Reaktionszeiten vernachlässigt werden. Die Datenerfassung erfolgte in Echtzeit mit einer Zykluszeit von zwei Millisekunden, wodurch sehr genau die Reaktionszeit gemessen werden konnte. Die Ergebnisse der Messungen wurden dann zusammen mit den Experimentdaten (Abfolge der Trials) in einer Datei gespeichert und weiteren statistischen Verarbeitungsschritten zur Verfügung gestellt.

Ergebnisse

Die deskriptive Analyse zeigt, dass sich die Bearbeitungszeiten der beiden Aufgabentypen auf den zwei Tiefenebenen nicht gleichermaßen unterscheiden, denn die Numerische Intelligenzaufgabe wird auf der hinteren Bildebene schneller bearbeitet, während sich bei der Detektionsaufgabe keine Unterschiede zeigen (siehe Abbildung 2):

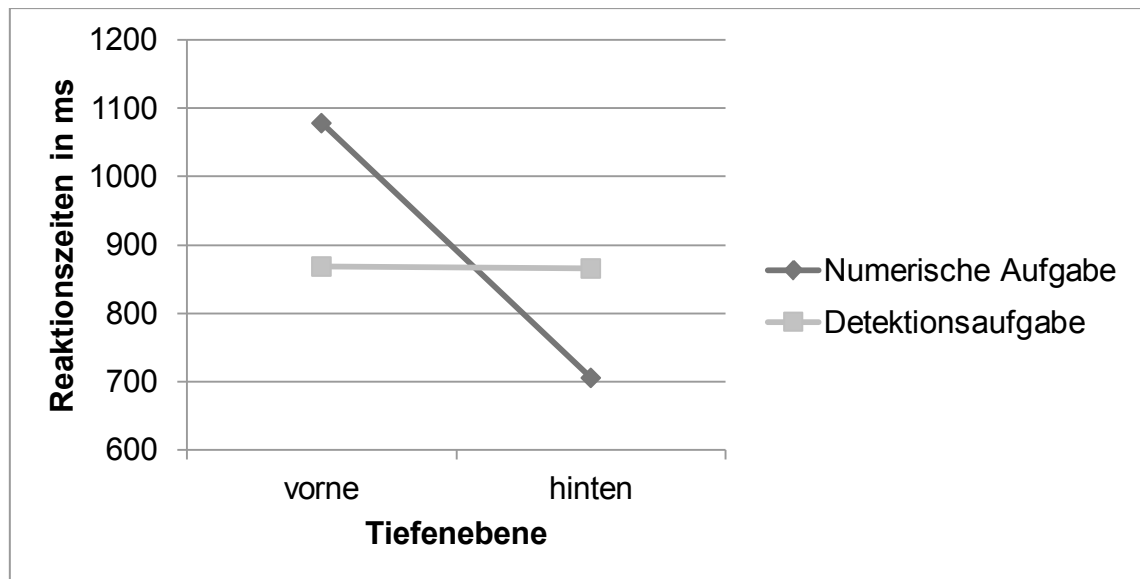


Abbildung 11: Ergebnisse: Reaktionszeiten der beiden Aufgabentypen auf den beiden Tiefenebenen

Die hier dargestellten Werte sind die Mittelwerte der Experimental Trials über alle Versuchspersonen. Anschließend wurden die Daten inferenzstatistisch analysiert, wobei zur genaueren Betrachtung des Effekts der Tiefenebene t-Tests pro Aufgabentyp durchgeführt wurden. Bezüglich der Reaktionszeiten der einfachen Objektdetektionsaufgabe (840 Millisekunden vorne/ 846 Millisekunden hinten) konnte kein signifikanter Effekt nachgewiesen werden ($p = 0,41$). Die mittleren Reaktionszeiten der numerischen Intelligenzaufgabe wurden hingegen mit 988 Millisekunden auf der hinteren Ebene signifikant schneller bearbeitet, als auf der vorderen Ebene (1047 Millisekunden): $t(21) = 10,74$; $p < 0,001$.

Diskussion

Die Studienergebnisse lassen einen Einfluss des Konvergenz-Akkommodations-Konflikts auf der vorderen Tiefenebene vermuten.

In der vorderen Tiefenebene, die sich im Abstand einer Armlänge (ca. 65 Zentimeter) vom Probanden befand, kam dieser Konflikt wohl voll zum Tragen, da Konvergenz und Akkommodation bis zu einer Distanz von bis zu circa drei Metern als Tiefenhinweise dienen (Cutting & Vishton, 1995; Leibowitz et al., 1972). Somit könnten das Wahrnehmen und Verarbeiten der dort präsentierten Information durch den Konvergenz-Akkommodations-Konflikt, d.h. die Entkopplung von Konvergenz und Akkommodation, erschwert worden sein. Die Augen fokussierten bzw. akkommodierten in allen Durchgängen stets auf eine Distanz von ca. zwei Metern, bei Darstellung der Objekte in der vorderen Ebene konvergierten die Augen jedoch lediglich auf eine Distanz von ca. 65 Zentimetern. Es gab somit sich widersprechende Tiefenhinweise.

Die hintere Tiefenebene lag mit einem Abstand von ca. 465 Zentimetern (Armlänge + vier Meter) hingegen außerhalb des Bereichs, in dem Konvergenz als Tie-

fenhinweis dient. Somit ist es wahrscheinlich, dass der Konvergenz-Akkommodations-Konflikt in dieser Ebene keinen Einfluss auf die Wahrnehmung hatte, da nur die Akkommodation (in ca. zwei Metern Entfernung), nicht mehr jedoch die Konvergenz, als Tiefenhinweis herangezogen wurde. Es kam somit zu keinem Widerspruch der beiden okulomotorischen Hinweise. Dies könnte erklären, warum die numerische Intelligenzaufgabe in der vorderen Ebene im Mittel langsamer bearbeitet wurde als in der hinteren Ebene.

Die Tatsache, dass die einfache Objektdetektionsaufgabe keinen derartigen Effekt aufweist, könnte darauf zurückzuführen sein, dass das Objekt hierbei nicht zwingend fixiert, sondern lediglich im Blickfeld bemerkt werden musste. Dieser Unterschied der Aufgabentypen könnte somit für die unterschiedliche Reaktion auf den primär in der vorderen Ebene vorherrschenden Konvergenz-Akkommodationskonflikt sein.

Die Konsequenz für die Validierung von Cockpitsystemen, welche durch virtuelle Modelle abgesichert werden, ist daher, dass die Informationen nicht näher dargestellt werden sollten, als die Projektionsfläche selbst entfernt ist. So kann der Einfluss des Konvergenz-Akkommodations-Konflikts vermindert werden und von einer Übertragung der Validierungsergebnisse ins Reale ausgegangen werden.

Quellen

- Cutting, J. E., Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers (Hrsg.), *Handbook of perception and cognition: Perception of space and motion*. New York: Academic Press, 69-117.
- Dittrich, E. und Israel, J. H. (2012). Arrangement of Product Data in CAVE Systems. Poster VisWeek 2012, Seattle
- Hoffman, D. M., Girshick, A. R., Akeley, K., Banks, M. S. (2008). Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3):33, 1-30.
- Israel, J. H., Belaifa, O., Gispén, A., & Stark, R. (2011). An Object-centric Interaction Framework for Tangible Interfaces in Virtual Environments. *Proceedings of Fifth international conference on Tangible, embedded, and embodied interaction ACM TEI'11*, Funchal, Portugal, ACM Press (pp. 325–332).
- Leibowitz, H. W., Shina, K., Hennessy, R. T. (1972). Oculomotor adjustments and size constancy. *Perception and Psychophysics*, 12, 497-500.
- Lichtenstein, A. (2009). Task-Specificity and Resource Allocation in Information Perception in Three-Dimensional Space. *Proceedings of HCI 2009 Conference*, 1-4 September 2009, Cambridge, UK. 184-187.
- Previc, F. H. (1998). The Neuropsychology of 3-D Space. *Psychological Bulletin*, 124 (2), 123-164.

ATTENDO: Methode zur Erfassung visueller Aufmerksamkeitsverteilung

Axel Hillebrand¹

Schlüsselwörter: Visuelle Aufmerksamkeit, Messmethode, Cognitive Tunnelling, Peripheral Vision Horizon Display, ATTENDO

Zusammenfassung

ATTENDO (lateinisch für *ich richte meine Aufmerksamkeit auf etwas*) ist eine Methode zur Erfassung der visuellen Aufmerksamkeitsverteilung (VAV) bei der Nutzung komplexer Displays, wie sie in Flugzeugcockpits, in der Luftraumüberwachung oder zur Steuerung von Unmanned Aerial Vehicles (UAVs) eingesetzt werden. Sie wurde bei EADS Innovation Works entwickelt.

Diese Methode soll Entwickler unterstützen, Displaydesign-Varianten zu identifizieren, die aufgrund ihrer Form der Informationsdarbietung zu einer ungeeigneten Verteilung von visueller Aufmerksamkeit führen. Diese wiederum kann zur Folge haben, dass für die durchzuführende Aufgabe oder Mission wichtige Informationen vom Bediener nicht wahrgenommen werden. In diesem Beitrag werden die Methode sowie ihr Einsatz in zwei Vorstudien detailliert vorgestellt. Außerdem wird der Stand der Planung für eine Hauptuntersuchung erläutert.

Einführung

Gibt es in einem Bereich eines Displays eine zu große Anzahl von eingeblendeten Informationen, kann dies zum Effekt des Cognitive Tunnelling führen. In diesem Fall werden andere Bereiche des Displays weniger intensiv beobachtet, und wichtige Informationen für die Sicherheit und die Effizienz eines Fluges oder einer Mission werden nicht beachtet. Basierend auf dem Secondary-Task-Paradigma erfasst ATTENDO die VAV mit Hilfe einer Detektionsaufgabe. Systematisch über das gesamte Display verteilt, werden Targets in Form von Quadraten zeitlich und räumlich randomisiert für eine halbe Sekunde eingeblendet. Der User hat neben seiner Primäraufgabe des Landeanflugs oder der Steuerung eines UAV die Aufgabe, per Tastendruck so schnell wie möglich auf ein eingeblendetes Target zu reagieren. Die VAV wird anhand der Detektionsrate der Targets in den verschiedenen Bereichen eines Displays operationalisiert.

ATTENDO wurde bereits in zwei Untersuchungen eingesetzt: bei der Evaluierung von Head-up Displays (HUDs, Vorstudie 1) und als dreidimensionale Variante (ATTENDO^{3D}) in einer stereoskopischen Workbench (Vorstudie 2).

¹ EADS Innovation Works

Andere Methoden zur Erfassung der VAV nutzen die Detektionsrate von plötzlich auftauchenden Hindernissen, z.B. in Form von Flugzeugen (Runway Incursions), oder andere überraschende Ereignisse (z. B. Alexander et al., 2005; Strayer, 2003). Jedoch liefert diese Vorgehensweise nur in geringem Umfang Rohdaten, da ein Surprising Event per Definition demselben Probanden nicht häufiger präsentiert werden kann. Eine niedrige statistische Aussagekraft ist die Folge. Auch die Bewertung der Blickbewegung stellt kognitionspsychologisch gesehen keine vollwertige Alternative dar, da z. B. nicht zwischen visueller und non-visueller Aufmerksamkeit unterschieden werden kann (Wickens & McCarley, 2008).

Es ist eine Hauptuntersuchung geplant, in der mit Hilfe von ATTENDO die VAV beim Einsatz eines innovativen Cockpitkonzeptes getestet wird. Dieses Konzept beinhaltet ein komplexes HUD und ein Peripheral Vision Horizon Display (PVHD). In der hier verwendeten Version präsentiert Letzteres dem Piloten den künstlichen Horizont zusätzlich in den Säulen des Cockpitfensters und soll somit bei der gleichzeitigen Nutzung des HUDs dem Effekt des Cognitive Tunnelling entgegenwirken. ATTENDO wird bei der Evaluierung dieser Systemkombination nicht nur in der zweidimensionalen, sondern auch in der dreidimensionalen Variante eingesetzt. Erstmals kann somit die VAV bei Nutzung eines HUDs auch in der Tiefenebene erfasst werden. Die Versuche werden in einem stereoskopischen Display erfolgen.

Im Folgenden werden ATTENDO näher beschrieben sowie die beiden Vorstudien und der Planungsstand der Hauptuntersuchung erläutert.

ATTENDO

Die Idee von ATTENDO basiert auf dem Secondary-Task-Paradigma. Die Secondary Task wird hier nicht genutzt, um den Workload des Nutzers, sondern um die Verteilung der visuellen Aufmerksamkeit zu erfassen (Hillebrand et al., 2012; Hillebrand, 2013).

Im Display werden für eine halbe Sekunde Targets eingeblendet, auf die der Nutzer so schnell wie möglich per Tastendruck reagieren muss. Das Einblenden der Targets erfolgt zeitlich und räumlich randomisiert. Die Zeit zwischen den Targets beträgt 7–15 Sekunden. Die Positionen der Targets können systematisch über dem gesamten Display verteilt sein oder sich in nur für die jeweilige Fragestellung interessanten Bereichen des Displays befinden. Mehrere Target-Positionen können zu einem Bereich zusammengefasst werden.

Operationalisiert wird die Verteilung der visuellen Aufmerksamkeit über die Detektionsrate der Targets. Werden in einem Bereich mehr Targets detektiert als in einem anderen, ist die visuelle Aufmerksamkeit unterschiedlich verteilt.

ATTENDO wird als Software in die Simulation integriert (Hillebrand, 2012). Um in einer bestimmten Versuchsumgebung ATTENDO nutzen zu können, muss

demnach u. U. ein nicht unerheblicher Programmieraufwand betrieben werden, der die Kosten der jeweiligen ATTENDO-Applikation in die Höhe treiben kann. Bei der hier vorgestellten Vorstudie 1 wurde ein Plug-in für den kommerziellen Flugsimulator X-Plane erstellt, das an jedem Rechner eingesetzt werden kann. Für Vorstudie 2 wurde eine individuelle Software-Lösung für ATTENDO^{3D} erstellt, die nur in Verbindung mit der verwendeten Workbench genutzt werden kann, und für die Hauptuntersuchung wurde eine Kombination aus einem Plug-in für X-plane sowie einer individuellen Lösung zur stereoskopischen Darstellung programmiert. Der Programmieraufwand für jede dieser ATTENDO-Applikationen lag bei 10-14 Manntagen.

Vorstudie 1

Zum ersten Mal angewendet wurde ATTENDO in einer Studie, in der unterschiedliche Head-up-Displays bezüglich der bei ihrem Einsatz gemessenen VAV evaluiert wurden (Hillebrand et al., 2012). Ziel der Untersuchung war es zu überprüfen, ob ATTENDO in der Lage ist, Unterschiede in der visuellen Aufmerksamkeitsverteilung zu detektieren. Das HUD wurde als Anwendungsbeispiel ausgewählt, da hier Beeinträchtigungen bezüglich der VAV bereits aus der Literatur bekannt sind (siehe z. B. Fadden et al., 1998; Foyle et al., 1991; Thomas & Wickens, 2001).

In der Studie wurden die Targets in drei Bereichen mit insgesamt 29 Targetpositionen eingeblendet: (1) im Zentrum, (2) am Rand und (3) an der Peripherie des HUDs. Die folgende Abbildung (Abb. 1) zeigt diese drei Bereiche, die mehrere Target-Positionen zusammenfassen. Sie sind zur besseren Anschauung grau unterlegt. Anders als im Versuchsablauf werden alle Targets gleichzeitig angezeigt.

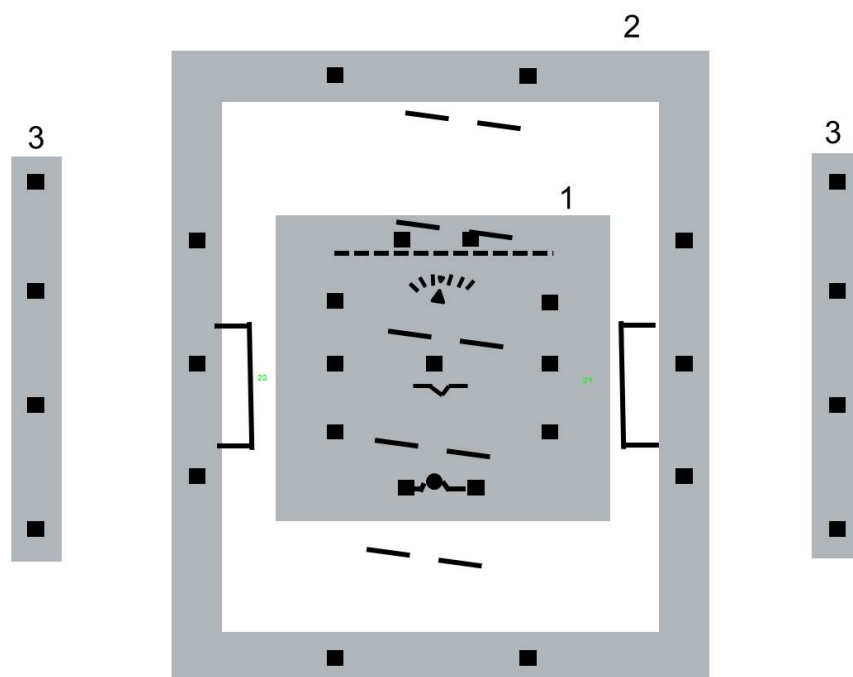


Abb.1: HUD- Symbologie mit ATTENDO-Target-Positionen (Hillebrand et al., 2102)

An der Studie nahmen zwölf lizenzierte Piloten teil. Sie fand statt in einer Low-Fidelity-Versuchsumgebung bestehend aus einem Monitor sowie handelsüblichem Joystick und Throttle. Als Flugsimulation diente X-Plane mit ATTENDO-Plug-in. Die Hauptaufgabe bestand darin, insgesamt sechs Landeanflüge mit Hilfe eines HUDs zu absolvieren und dabei so nah wie möglich am idealen Verlauf der Anflugbahn entlangzufliegen. Head-down-Instrumente standen dafür nicht zur Verfügung. Zusätzlich sollten die Piloten bei Detektion eines Targets so schnell wie möglich eine Taste am Joystick drücken.

Verglichen wurden zwei HUD-Versionen, die sich in der Anzahl der präsentierten Informationen unterschieden: ein rudimentäres und ein komplexes HUD. Beide HUD-Versionen beinhalteten alle für einen Landeanflug notwendigen Informationen. Das komplexe HUD präsentierte den Piloten darüber hinaus eine Tunnel-in-the-Sky-Funktion, einen Flight Director und ein Aircraft Reference Symbol.

Aus den zwei HUD-Versionen und den drei Bereichen der Target-Positionen ergab sich ein 2x3-Design. Die statistische Auswertung zeigte, dass Targets, die im Zentrum des HUDs (1) und am Rand des HUDs (2) präsentiert werden, häufiger detektiert werden, als Targets an der Peripherie (3). Es gab zwischen den HUD-Versionen keinen Unterschied bezüglich der Detektionsrate, was mit einem zu geringen Ausmaß an Variationsmöglichkeiten zusammenhängen könnte, die X-Plane bereitstellt, um unterschiedliche HUD-Variationen zu erstellen.

Die signifikanten Unterschiede zwischen den Bereichen der Target-Positionen zeigen, dass ATTENDO durchaus in der Lage ist, sich auf die Aufmerksamkeitsverteilung auswirkende Zustände zu reproduzieren.

Vorstudie 2

In einer zweiten Vorstudie wurde ATTENDO^{3D} zur Erfassung der visueller Aufmerksamkeitsverteilung im dreidimensionalen Raum eingesetzt (Michalczyk et al., 2013). Im Folgenden werden nur diejenigen Inhalte der recht umfassenden Studie aufgezeigt, die für den Einsatz von ATTENDO^{3D} relevant sind.

Ziel von Vorstudie 2 war es u. a. zu überprüfen, ob ein Einsatz von ATTENDO auch im dreidimensionalen Raum möglich ist. Zu diesem Zweck wurde eine Applikation von ATTENDO programmiert, die in einer 3D-Workbench zum Einsatz kommen kann. Diese 3D-Workbench bietet dem Benutzer in einem Areal von etwa 80cm x 60cm x 60cm einen dreidimensionalen Raum dar.

In dem dreidimensionalen Raum wurden 27 Target platziert. Abbildung 2 zeigt deren Verteilung. Es gibt neun Target-Positionen in einer vorderen Ebene, neun in einer mittleren und neun in einer hinteren.

In Vorstudie 1 bindet das HUD die Aufmerksamkeit in einem bestimmten Areal des Displays. Da für die 3D-Workbench keine praxisbezogene Anwendung verfügbar war, wurde nach einer Aufgabe gesucht, die die visuelle Aufmerksamkeit auf sich zieht und in einem lokal begrenzten Areal des Displays präsentiert werden

kann. Die Entscheidung fiel auf den Stroop Test. Er ist zudem eine Aufgabe, die auch von nahezu jedermann schnell erlernbar ist. Dem Probanden werden Farbwörter nacheinander gezeigt. Der Proband hat zur Aufgabe, die Tintenfarbe zu benennen, die allerdings nicht der Bedeutung des Wortes entspricht. Die gesagte Tintenfarbe wurde in der vorliegenden Studie auf Tonband aufgenommen. Es gab für den Stroop-Test zwei Positionen (ST_{P1} und ST_{P2} , siehe Abb.2), an denen die Wörter nacheinander eingeblendet wurden: in der vorderen Ebene unten links und in der hinteren Ebene oben rechts. Der Blickwinkel der Probanden, die vor der Workbench saßen, war schräg von oben (siehe Abb. 2).

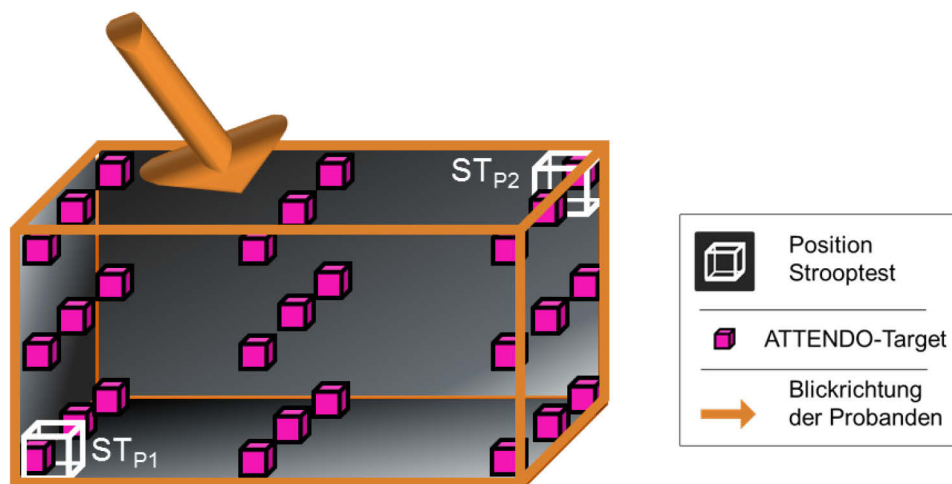


Abb. 2: Target-Positionen, Positionen des Stroop-Tests und Blickwinkel der Probanden

An der Studie nahmen 24 Probanden teil, und sie beinhaltete zwei Durchgänge: für jede Position des Stroop-Tests einen Durchgang. Pro Durchgang wurden 108 Wörter präsentiert. Wie in Vorstudie 1 wurden die Targets räumlich und zeitlich (alle 7-15 Sekunden) randomisiert für eine halbe Sekunde eingeblendet.

Für die statistische Auswertung wurden die Targets wie folgt gruppiert: Eine Gruppe bildeten die Target-Positionen, die sich direkt an der Position des Stroop-Tests befanden. Dabei handelte es sich um acht Target-Positionen. Die andere Gruppe bildeten die übrigen 19 Target-Positionen. Abbildung 3 verdeutlicht diese Gruppen für eine Position des Stroop-Tests (ST_{P1})

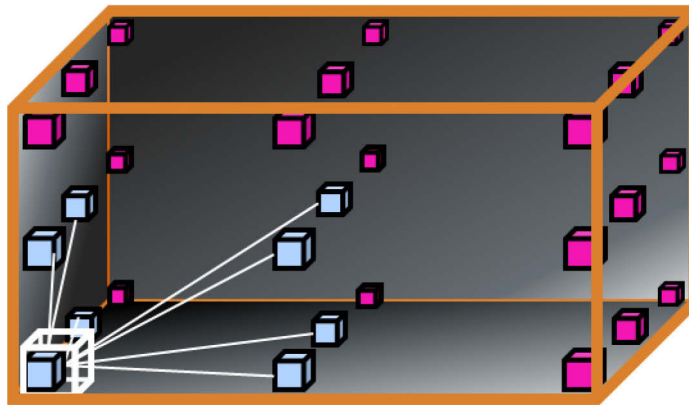


Abb. 3: Gruppierung der Target-Positionen ST_{P1}

Es zeigte sich für beiden Positionen des Stroop-Tests (ST_{P1} und ST_{P2}), dass die Targets, die in der unmittelbaren Nähe eingeblendet werden, signifikant häufiger detektiert werden. Dieses Ergebnis lässt darauf schließen, dass auch die ATTENDO-Applikation für dreidimensionale Räume in der Lage ist, die Verteilung der Aufmerksamkeit zu erfassen.

Planungsstand der Hauptstudie

Die Ergebnisse der beiden Vorstudien sollen für eine Hauptstudie genutzt werden. Wie in Vorstudie 1 hat diese einen Bezug zur Luftfahrt. Es soll ein Display evaluiert werden, das dem Cognitive Tunnelling bei der Nutzung eines HUDs entgegenwirkt. Bei diesem System handelt sich um das Peripheral Vision Horizon Display (PVHD). Dieses System präsentiert dem Piloten in den Säulen des Cockpitfensters den künstlichen Horizont, der die Lage des Flugzeugs im Raum anzeigt. Somit steht dem Piloten der künstliche Horizont dreimal zu Verfügung: im HUD, also im Zentrum des Frontfensters, sowie links und rechts davon in den Säulen des Cockpitfensters. Diese zusätzlichen Displays sollen die ungünstige Zentrierung der visuellen Aufmerksamkeit aufbrechen und deren Verteilung gleichmäßiger über die gesamte vordere Außensicht lenken.

ATTENDO soll bei dieser Evaluation wie folgt zum Einsatz kommen: Wie in Vorstudie 1 wird es Target-Positionen im Zentrum des HUD (zwei Positionen) und an der Peripherie (zwei links vom HUD und zwei rechts davon) geben. Damit wird die Aufmerksamkeitsverteilung in x- und y-Richtung erfasst. In Anlehnung an Vorstudie 2 wird zusätzlich die Aufmerksamkeit in z-Richtung, also der Tiefenebene, erfasst. Dazu wird es nicht nur eine Ebene dieser sechs Target-Positionen in Nähe der Cockpit-Scheibe geben, sondern eine weitere, die sich mehrere Meter vor dem Flugzeug befindet. Daraus ergeben sich zwölf Target-Positionen.

Realisiert wird dieser Versuchsaufbau in einem Virtual-Reality-Labor an einer Powerwall. Mit Hilfe von zwei Projektoren und passiver Filterung wird ein stereoskopisches Bild erzeugt. Wie in Vorstudie 1 wird die Flugaufgabe mit der Flugsimulation X-Plane erzeugt, und die Targets werden mit einem entsprechend für diesen Zweck programmierten Plug-in dreidimensional eingeblendet. Die Flugaufga-

be wird sich von der in Vorstudie 1 unterscheiden. So werden keine Landeanflüge gemacht, sondern es wird mit Hilfe des Flight Directors im HUD ein kurvenreicher Parcours abgeflogen, wobei eine vorgegebene Geschwindigkeit eingehalten werden soll. Da für diese Trackingaufgaben kein Fachwissen erforderlich ist, ist es möglich, sowohl lizenzierte Piloten als auch erfahrene Spieler von Flugsimulationen, sogenannte Simmer, als Probanden heranzuziehen. Dadurch können die Kosten für die Studie erheblich gesenkt werden.

Bei der Hauptstudie wird es zudem eine Neuerung bezüglich ATTENDO geben. Es wird keine starre Target-Position mehr geben. Für die Targets wird es ein begrenztes Feld geben, in dem das Target an unterschiedlichen Positionen erscheint. Da jedes Target viermal präsentiert wird, soll durch diese Neuerung die Vorhersehbarkeit für das Erscheinen eines Targets reduziert werden.

Mit Hilfe der Hauptstudie sollen zwei Fragen beantwortet werden: Ist das PVHD in der Lage, die Konzentration der Aufmerksamkeit im Zentrum zu reduzieren? Und: Gibt es Unterschiede bezüglich der Verteilung der visuellen Aufmerksamkeit bei der Nutzung eines HUD nicht nur bezüglich Zentrum und Peripherie, sondern auch in der dritten Ebene, der Tiefenebene?

Literatur

- Alexander, A. L., Wickens, C. D. & Hardy, T. J. (2005). Synthetic Vision Systems: The Effect of Guidance Symbolology, Display Size, and Field of View. *Human Factors*, 47(4), 693-707.
- Fadden, S., Ververs, P. M. & Wickens, C. D. (1998). Costs and Benefits of Head-Up Display Use: A Meta-Analytic Approach. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 42 (1), 16-20.
- Foyle, D. ., Sanford, B. D. & McCann, R. S. (1991). Attentional Issues in Superimposed Flight Symbolology. *Proceedings of the Sixth International Symposium on Aviation Psychology*, 577-582.
- Hillebrand, A. (2012). Methoden zur Evaluierung des User-Interfaces großer Displays. In M. Grandt & S. Schmerwitz (Hrsg.), *Fortschrittliche Anzeigesysteme für die Fahrzeug- und Prozessführung* (S. 273–282). Bonn: DGLR.
- Hillebrand, A., Wahrenberg, E. & Manzey, D. (2012). *New Method to assess Pilots' Allocation of Visual Attention Using a Head-up Display*. In D. de Waard, K. Brookhuis, F. Dehais, C. Weikert, S. Röttger, D. Manzey et al. (Hrsg.), *Human Factors: a view from an integrative perspective*. *Proceedings of HFES Europe Chapter Conference Toulouse*. Verfügbar unter: <http://hfes-europe.org>.
- Hillebrand, A. (2013). ATTENDO: Method for the Assessment of Visual Attention Allocation in Two-dimensional Spaces. *Proceedings of 57th Annual Meeting of HFES* (accepted).

- Michalczyk, F., Hillebrand, A. & Thüring, M. (2013). *Domain Independent Visual Attention Assessment in Stereoscopic Displays*. In: D. de Waard (Hrsg.), Human Factors: sustainable life and mobility. Proceedings of HFES Europe Chapter Conference Toulouse (accepted).
- Strayer, D.L., Drews, F.A. & Johnston, W.A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied*, (9), 23-32.
- Thomas, L. C. & Wickens, C. D. (2001). Visual Displays and Cognitive Tunneling: Frames of Reference Effects on Spatial Judgments and Change Detection. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 45 (4), 336-340.
- Wickens, C. D. & McCarley, J. S. (2008). *Applied Attention Theory*. Boca Raton: CRC Press.

Einfluss der Audiomodalität auf die Wahrnehmung und Qualitätsbeurteilung (audio-)visueller Stimuli

Sebastian Arndt¹, Juliane Bürglen, Jan-Niklas Antons,
Robert Schleicher und Sebastian Möller

Schlüsselwörter: Videoqualität, Elektroenzephalogramm, Mean Opinion Score (MOS), audiovisuell, P300

Zusammenfassung

Bei Onlinevideodiensten ist es wichtig, eine Balance zwischen bestmöglicher audiovisueller Qualität und kleinstmöglicher Verzögerung des Videos zu finden. Um flüssige (Live-)Streams zu gewährleisten, muss bei eingeschränkter Bandbreite die zu übertragende Datenmenge reduziert werden. Hierbei ist von Belang, ob und wie Veränderungen in der Audiomodalität die visuelle Qualitätswahrnehmung beeinflussen, und andersherum. In der Modalität, von der weniger Einfluss ausgeht, könnte Bandbreite gespart werden, ohne dass die wahrgenommene Qualität reduziert wird. Die vorliegende Studie befasste sich mit einem Zwischenschritt dieser Frage, indem Clips in verschiedener Videoqualität mit und ohne Audiospur gezeigt wurden. Mit der verwendeten EEG-Methode konnten bislang die vorbewusste Wahrnehmung sowohl von Audio- als auch von Videoqualität separat untersucht werden; in dieser Studie wurden beide Modalitäten erstmals kombiniert. Bei rein visueller Präsentation wurde die visuelle Qualität subjektiv schlechter bewertet als mit Audio, im EEG konnten qualitätsabhängige P300-Komponenten gefunden werden. Es konnte gezeigt werden, dass das Paradigma auf audiovisuelle Stimuli übertragbar ist.

Messung von Videoqualität

Da Qualität subjektiv ist, also in der Wahrnehmung des Betrachters liegt und nicht an sich physikalisch messbar ist, werden zur Erhebung des Qualitätsempfindens subjektive Messmethoden verwendet. Kontexteinflüsse sollen hierbei durch möglichst weitgehende Standardisierungen minimiert werden (siehe ITU-R Rec. BT.500-12, 2009). Im Falle einer Bewertung ohne Referenz wird dazu die "Absolute Category Rating" (ACR) Skala verwendet, welche oftmals mit fünf diskreten Abstufungen mit den Labels "Ausgezeichnet - Gut - Ordentlich - Dürftig - Schlecht" benutzt wird. Mit der Methode des Mean Opinion Score (MOS) werden die von den Probanden erhaltenen Werte auf dieser Skala gemittelt. Falls ein Referenzstimulus mit verwendet wird, soll der betreffenden Stimulus in Relation zum ersten bewertet werden, ein Beispiel dafür ist das "Degradation Category Rating" (DCR).

¹ Alle Autoren: Technische Universität Berlin, Quality und Usability Lab

Qualitätsmessung mit EEG

Bei subjektiven Methoden findet zwischen der Sinneswahrnehmung und dem Urteil immer eine Interpretation statt, die abhängig von Kontext und Betrachter ist. Um diese Bestandteile genauer zu untersuchen, kann man bereits beim Wahrnehmungsvorgang im Gehirn ansetzen, um unter anderem vorbewusste, neuronale Reaktionen auf Qualitätsunterschiede zu messen.

Auf der Suche nach nichtinvasiven Methoden hierfür ergaben Studien von z.B. Antons et al. (2012), dass vorbewusste Qualitätswahrnehmung von auditiven Stimuli per EEG messbar ist. Arndt et al. (2011) konnten dies auch für Videomaterial (unimodal) zeigen. Es wurden Komponenten der Ereigniskorrelierten Potenziale (EKP), unter anderem die P300, extrahiert deren Ausprägung mit der jeweiligen Qualitätsabstufung der Stimuli variierte. In dieser Studie wird untersucht, ob ein solches Paradigma auch bei audiovisuellen Stimuli anwendbar ist und ob eine Videoqualitäts-abhängige P300-Komponente, eine Positivierung im EKP, durch reine Anwesenheit einer Audiospur beeinflusst wird.

Methoden

Zehn Probanden im Alter von 22 bis 37 Jahren ($MW=28$, $SD = 5.2$) nahmen an der Studie teil. Alle Teilnehmer waren normalsichtig (ggf. mit Sehhilfe) und hatten kein beeinträchtigtes Hörvermögen.

Das EEG-System der Firma Brain Products GmbH, Garching, wurde mit 64 Elektroden, nach dem 10-20 Standardsystem (Jasper, 1958), verwendet. Zusätzlich wurde ein Elektrookulogramm (EOG) aufgezeichnet. Die Videos (Bildgröße 35.6x20cm) wurden im Abstand von 80 cm auf einem 24-Zoll-Monitor präsentiert, die Audiospur wurde über ohrumschließende Kopfhörer mit 65 dB abgespielt. Die Studie wurde in einem ITU-T Rec. P.910 (2008) -konformen Raum mit grauen Vorhängen und künstlichem Tageslicht durchgeführt. Für die Aufgabe der Probanden während der EEG-Messung stand eine Tastatur mit zwei beschrifteten Tasten zur Verfügung ("Unterschied", "kein Unterschied").

Stimuli

Die Stimuli wurden aus kurzen Videoclips zusammengesetzt, welche die untere Gesichtshälfte einer Frau zeigten, während diese die Silbe "pa" ausspricht. Die Auswahl begründete sich durch die beabsichtigte Vergleichbarkeit mit bereits publizierten EEG-Audioqualitätsstudien (siehe Antons et al., 2012) und ähnelte stark denen der vorangegangenen EEG-Videoqualitätsstudie (siehe Arndt et al., 2011).

Aus der Originalaufnahme mit hoher Audio- und Videoqualität wurden 20 verschiedene Versionen erstellt, indem jeweils mit und ohne Ton eine Videostörung in zehn Stärken eingebaut wurde. Die Störung vom Typ "Blockiness" sollte der aus dem Alltag bekannten "Verpixelung" entsprechen und wurde mit dem in ITU-T P.930 (1996) beschriebenen Algorithmus erstellt. Ein Durchgang bestand aus ei-

nem Clip-Paar: Jeweils dem Original-Videoclip mit Ton, gefolgt von einem möglicherweise gestörtem Clip mit oder ohne Ton. Zwischen den beiden wurde ein kurzes Morphing (fließender Übergang) eingefügt, damit kein sichtbarer Schnitt die EEG-Aufnahme beeinflusste.

Während der Kalibrationsphase war es das Ziel, zwei Störungsstufen zu finden, deren Erkennungsrate bei ca. 33% (LQI) und ca. 66% liegen sollte (LQII). Damit sollte sichergestellt werden, dass die Aufgabe für alle Probanden etwa gleich schwierig war. Weiterhin wurden im Hauptversuch noch die unveränderte Videoqualität (HQ) und die schlechteste Qualität (LQmax) verwendet.

Hauptexperiment

Das 4x2-Design bestand aus den Innersubjekt-Faktoren Videostörung (Stufen HQ, LQI, LQII, LQmax) und Audiomodus (Ton an: AV, Ton aus: V). Im EEG-Hauptexperiment war es Aufgabe der Probanden, den zweiten Teil des Clip-Paares mit dem ersten hinsichtlich der Videoqualität zu vergleichen (Two-alternative forced choice task). Der erste Teil war immer ungestört, der zweite konnte gestört sein. Die Probanden wurden instruiert, nach Stimuluspräsentation eine der beiden Tasten zu drücken; je nachdem, ob sie einen Unterschied in der Videoqualität wahrgenommen hatten oder nicht. Der Audiomodus sollte ignoriert werden. Jede der acht kombinierten Bedingungen wurde, in randomisierter Reihenfolge, 100 mal gezeigt.

Subjektive Qualitätsbewertung

Im Teil der subjektiven Qualitätsbeurteilung (ohne EEG) wurden die Clip-Paare in allen 20 Versionen gezeigt (Innersubjekt-Faktoren: Videostörung (Zehn Videostörungs-Stufen) und Audiomodus (Ton an/aus)). Das Qualitätsurteil der Teilnehmer (wieder nur in Bezug auf Videoqualität) wurde auf einer siebenstufigen kontinuierlichen ACR-Skala erfasst, mit entsprechenden Beschriftungen, die eine Erweiterung der fünfstufigen Standard-MOS-Skala darstellte und auf den beiden Extrema keine Labels trug (ähnlich wie Bodden & Jekosch, 1996, zitiert nach Möller, 2010, S.44f).

Datenaufbereitung

Für die Verarbeitung der EEG-Daten und statistische Analyse wurden die Matlab-Toolbox EEGLab (Delorme & Makeig, 2004) sowie SPSS verwendet. Die EEG-Daten wurden mit 1000 Hz aufgenommen, auf 200 Hz reduziert, anschließend wurde ein Bandpassfilter von 0.1-45 Hz angewendet(nach Duncan et al., 2009). Epochen wurden von 200ms vor bis 1000ms nach Stimulusbeginn definiert und Baseline-korrigiert (-200ms bis 0 ms).

Ergebnisse: Erkennungsraten

Bedingt durch niedrige Erkennungsraten in der Bedingung LQI bei allen Probanden, sowie in der Bedingung LQII bei vier Probanden wurden diese nicht in der Varianzanalyse der Verhaltens- und EEG-Daten verwendet, da nicht genügend auswertbare Durchgänge vorlagen (siehe Duncan et al., 2009).

Für die restlichen Bedingungen und Probanden (drei Störungsstufen, zwei Audiomodi) wurde für die Erkennungsraten eine zweifaktorielle Varianzanalyse mit Messwiederholung berechnet. Hierbei war der Haupteffekt der Videostörungsstufe signifikant ($F(2, 10) = 262.11, p < 0.01$), jedoch nicht der des Audiomodus oder die Interaktion. Bonferroni-korrigierte Paarvergleiche ergaben, dass jeweils die HQ-Bedingung sich von den übrigen unterschied: HQ-LQII: $p < 0.01$, HQ-LQmax: $p < 0.01$. Die Erkennungsraten für HQ und LQmax lagen in beiden Audiomodi wie erwartet bei 100% oder knapp darunter. Für LQII lagen sie im Bereich von 73-100%.

Ergebnisse: EEG-Daten

Um in den EKP-Daten eine mögliche P300-Komponente zu erkennen, wurden alle korrekt beantworteten Durchgänge der Bedingung LQmax denen der Bedingung HQ gegenübergestellt. In LQmax fand sich im Vergleich, eine deutliche Positivierung in den Potentialen zentraler Elektroden im Zeitbereich von etwa 300 bis 700ms, mit dem Maximum bei 400 bis 450ms.

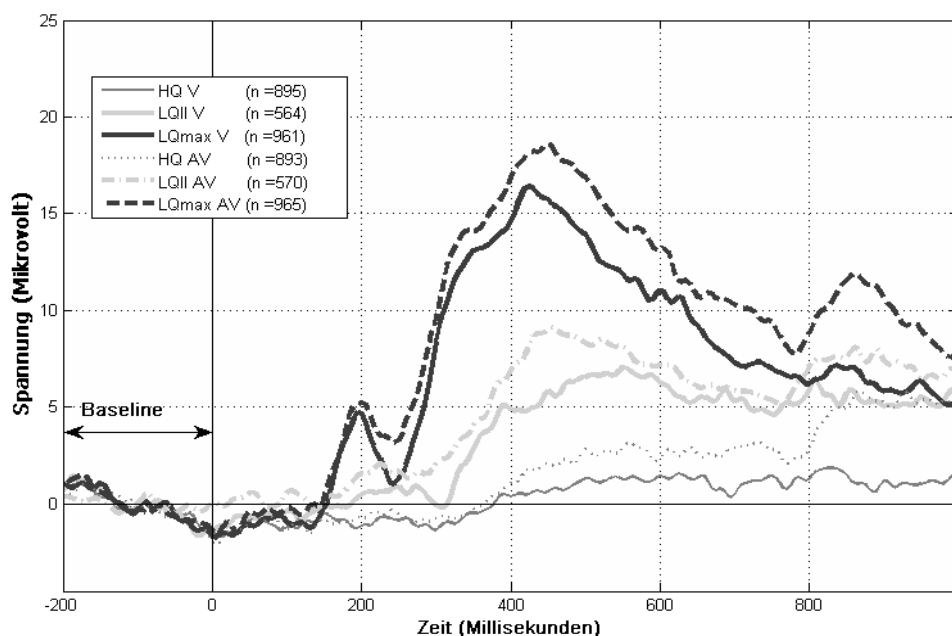


Abbildung 12: Grand Average der EKP-Daten für die Videostörungsstufen HQ, LQII und LQmax in schwarz, hellgrau und mittelgrau. Durchgezogene Linien bezeichnen die Bedingungen mit Audiospur. Die Anzahl der ausgewerteten (korrekt beantworteten) Durchgänge über alle Probanden ist mit n angegeben.

Für die EKP-Analyse wurden für Elektrode Cz der Grand Average für alle korrekt beantworteten Durchgänge berechnet, das heißt, für alle Hits (Taste „Unterschied“) in LQmax und LQII und alle Correct Rejections (Taste „kein Unterschied“) in HQ. Wie in Abbildung 1 dargestellt, zeigte sich die erwähnte Positivierung sehr ausgeprägt in LQmax, sowie weniger deutlich in LQII. In Durchgängen mit Audiospur (LQmax-AV und LQII-AV) schienen tendenziell die Positivierungen etwas größer auszufallen, außerdem trat eine kleinere Positivierung nach etwa 800ms auf.

Für die statistische P300-Analyse wurden die Daten von 300 bis 700ms nach Stimulus Onset ausgewählt, d.h. nach dem Beginn des zweiten Teils des Video-clip-Paares. In diesem Zeitfenster wurde der maximale Wert als Peak-Amplitude extrahiert. Eine messwiederholte zweifaktorielle Varianzanalyse der gemittelten Peak-Amplituden ergab keinen signifikanten Effekt für Audiomodus und Interaktion, jedoch einen signifikanten Unterschied für den Faktor Videostörung: $F(2, 10) = 48.68$, $p < 0.001$. Bonferroni-korrigierte Paarvergleiche ergaben, dass sich nur LQmax von den beiden anderen Störungsstufen unterschied (LQmax vs HQ: $p < 0.001$, LQmax vs. LQII: $p = 0.005$).

Auch die EKP-Positivierungen in den Audio-Durchgängen nach 800ms wurden durch eine ebensolche Varianzanalyse der gemittelten Peak-Amplituden getestet, hier zeigte sich jedoch kein signifikanter Effekt.

Ergebnisse: Subjektive Qualität

In den Daten der Subjektiven Bewertungen (alle 10 Störungsstufen mit den beiden Audiobedingungen) fielen die MOS-Werte für Durchgänge mit Audio in jeder Qualitätsstufe höher aus als ohne Audio. Wie erwartet waren die Beurteilungen besser, je weniger Blockiness das Video aufwies.

Zur Auswertung wurde ein zweifaktorieller messwiederholter Friedman-Test¹ mit den Faktoren Videostörung (10 Stufen) und Audiomodus (2 Stufen) durchgeführt. Hierbei zeigte sich ein signifikanter Effekt für den Audiomodus: $\chi^2(1, N = 8) = 24.87$, $p < .01$.

Diskussion

Für die Verhaltensdaten zeigte sich wie erwartet, dass die Erkennungsraten bei größerer visueller Störung höher waren, es gab jedoch nur signifikante Unterschiede zwischen LQmax und den beiden anderen Stufen (LQII, HQ). Bei LQII waren die Erkennungsraten generell höher als ursprünglich intendiert. Das Ziel, die Aufgabe für alle Probanden ähnlich schwierig zu machen, wurde hier aber größtenteils erreicht.

¹ Durch Speicherfehler konnten nur die Daten von 8 von 10 Versuchspersonen für die MOS-Analyse verwendet werden.

Die EKP-Analysen ergaben wie erwartet, dass eine störungsabhängige P300-Komponente gefunden werden konnte. Hier unterschied sich wieder nur LQmax signifikant von den anderen visuellen Qualitätsstufen in der Amplitude. Bei Anwesenheit der Audiospur fanden sich deskriptiv höhere gemittelte Peak-Amplituden, jedoch nicht signifikant. Prinzipiell konnte mit dieser Studie die Frage nach der Beeinflussung der Störungswahrnehmung durch Anwesenheit einer Audiospur nicht beantwortet werden.

Jedoch wurde ein zweiter nicht signifikanter Anstieg in den EKP nach 800ms gefunden, welcher nur in der Bedingung mit Audio vorkam. Die Ursache könnte in der Tatsache liegen, dass in der Audiospur der hörbare „pa“-Laut erst ab etwa 600ms nach dem Video-Stimulus-Onset einsetzte. Dies ist ein ähnlicher zeitlicher Abstand wie der der beiden auffälligen EKP-Positivierungen. Es kann angenommen werden, dass die Auffälligkeit eine P300-Komponente in Abhängigkeit des Audiomodus darstellt. Damit könnte zusammenhängen, dass die vorher untersuchten P300-Amplituden sich im Faktor Audio nicht signifikant unterschieden.

In folgenden Studien sollte darauf geachtet werden, dass der Stimulus Onset in beiden Modalitäten gleichzeitig stattfindet. Man kann erwarten, dass dann eine Beeinflussung der P300 deutlich wird.

Die Subjektiven Qualitätsbeurteilungen fielen erwartungsgemäß generell besser aus, je weniger gestört die Videos waren. Die MOS-Werte waren tendenziell für Clips mit Audiospur höher.

Es stellt sich für weiterführende Studien die Frage, ob Unterschiede sowohl in den subjektiven Bewertungen als auch in den EKP-Daten durch die reine Anwesenheit der Audiospur zustande kamen, oder durch die hohe Audioqualität bedingt sind und bei einer Audiostörung anders ausfallen würden.

Zusammenfassend lässt sich schlussfolgern, dass dieses Paradigma erfolgreich auch mit bimodalen Stimuli verwendet werden kann, auch wenn zukünftig ein Experimentalparadigma mit einem synchronen Onset gewählt werden muss. In zukünftigen EEG-Studien soll zusätzlich zur Videoqualität auch die Audioqualität variiert werden sowie in der Subjektiven Qualitätserfassung nach visueller Qualität, auditiver Qualität und audiovisueller Qualität unterschieden werden. Die Fortführung der Studienreihe kann dazu beitragen, mehr über den gegenseitigen Einfluss von wahrgenommener Video- und Audioqualität zu erfahren. Dies ist sowohl von Interesse für die Neuropsychologie, als auch für die Wirtschaft im zunehmend wichtigen Bereich der Onlinevideodienste.

Literatur

Antons, J., Schleicher, R., Arndt, S., Möller, S., Porbadnigk, A. K., & Curio, G. (2012). Analyzing Speech Quality Perception Using Electro-Encephalography. *IEEE Journal of Selected Topics in Signal Processing*, 6(6), 721–731.

- Arndt, S., Antons, J., Schleicher, R., Möller, S., Scholler, S., & Curio, G. (2011). A physiological approach to determine video quality. *Multimedia (ISM), 2011 IEEE International Symposium on*, 518–523.
- Delorme, A. & Makeig, S. (2004). EEGLAB: an Open Source Toolbox for Analysis of Singletrial EEG Dynamics Including Independent Component Analysis. *Journal of Neuroscience Methods*, 134(1), 9–21.
- Jasper, H.H. (1958). The Ten-Twenty Electrode System of the International Federation. *Electroencephalography and Clinical Neurophysiology*, 10(2), 371–375.
- Duncan, C., Barry, R.J., Connolly, J.F., Fischer, C., Michie, P.T., Näätänen, R., ... Van Petten, C. (2009). Event-Related Potentials in Clinical Research: Guidelines for Eliciting, Recording, and Quantifying Mismatch Negativity, P300, and N400. *Clinical Neurophysiology*, 120(11), 1883–1908.
- International Telecommunication Union. (1996). Principles of a Reference Impairment System for Video (ITU-T Recommendation No. P.930). Geneva.
- International Telecommunication Union. (2008). Subjective Video Quality Assessment Methods for Multimedia Applications (ITU-T Recommendation No. P.910). Geneva.
- International Telecommunication Union. (2009). Methodology for the Subjective Assessment of the Quality of Television Pictures (ITU-R Recommendation No. BT.500-12). Geneva.
- Möller, S. (2010). Quality Engineering. Qualität kommunikationstechnischer Systeme. Heidelberg: Springer.

Situation Awareness in Autonomous Service Robots

Verena Nitsch¹

Keywords: Cognitive Modeling, Situation Awareness, Autonomous Robots

Abstract

Despite the increasing demand for service robots, considerable challenges remain to be solved before these robots can be ubiquitously employed in personal households. To this date, autonomously operating robots lack the flexibility to react appropriately in unforeseen situations. At the Human Factors Institute of the Universität der Bundeswehr München, a multi-disciplinary research group of psychologists, engineers and computer scientists aims to ascertain to which extent cognitive models of situation awareness may enhance a robot's ability to display adaptive behaviour in unforeseen situations.

Introduction

In 2011, about 2.5 million robots were sold for personal and domestic use (International Federation of Robotics, 2012). The International Federation of Robotics projects sales of over 15 million units for the period 2012 to 2015 with experts indicating that service robots have an innovation and market potential similar to that currently presented by industrial robots (Decker, et al., 2011). To this date, autonomously operating robots lack the flexibility to react appropriately in unforeseen situations as important perceptual and decision-making structures are still designed by programmers before the robot begins its task. This leads to fragmented abilities and brittle performance, in particular in unstructured and dynamic environments, which are prevalent in domestic contexts (Benjamin, Monaco, Lin, Funk, & Lyons, 2012). Hence, perhaps the greatest challenge in the engineering of autonomously operating robotic systems remains the instillation of flexibility in the robot's decision-making processes that would allow it to adapt to the demands presented by a dynamically changing environment at any given moment, whilst still pursuing a particular operative goal (as opposed to purely reflexive systems). This line of research can be subsumed under the label of cognitive robotics.

In an effort to tackle this problem of failing adaptability in robots, several approaches have surfaced in the different scientific disciplines that are currently active in the field of cognitive robotics. One school of thought, which derived from the computer sciences, focuses primarily on the effective harnessing of computational processing capabilities and the development of mathematical algorithms to create learning and hence adaptable systems. Within the paradigm of machine learning, numerous attempts were made to develop heuristics based on statistical

¹ Universität der Bundeswehr München, Institut für Arbeitswissenschaft, Werner-Heisenberg-Weg 39, 85577 Neubiberg; Email: verena.nitsch@unibw.de

properties of sensor signals and reinforcement learning in an effort to reduce processing time and computational power requirements that would lead to slow response times and likely ineffective behaviour in dynamically changing situations (e.g. Bagnell, Bradley, Silver, Sofman, & Stentz, 2010; Peters, Morimoto, & Tedrake, 2009; Bratko, 2010). Although considerable progress has been made in this domain in recent years, real-time learning and in particular adaptive behavior of robots in unforeseen situations remain a great challenge in robotics (Benjamin, Monaco, Lin, Funk, & Lyons, 2012). A different approach may be sought to complement these efforts. Cognitive scientists have tackled the issue of system adaptability by investigating human cognitive abilities that lead to adaptive behavior in such situations. Based on decades of experimental psychological work, numerous models have been proposed in this domain that specify human cognitive processes in dynamic situations (e.g. Baumann & Krems, 2009; Gonzales & Dutt, 2007; Durso & Sethumadhavan, 2008). As of yet, however, the evaluation of these cognitive models remains chiefly theoretical work; to our knowledge, a truly cognitive architecture has not yet been implemented in a robotic system (Benjamin, Monaco, Lin, Funk, & Lyons, 2012). Consequently, it is not sufficiently clear to which extent relevant models of human cognitive processes may be applicable to autonomously operating robots and improve performance of such systems. Systematic investigations of the issues which would surface in the implementation of such models in robots, are hence urgently required.

Situation awareness as a pivotal cognitive process in dynamic environments

Cognitive researchers have suggested that the basis for humans' decision-making ability in dynamic situations forms their situation awareness (SA), a psychological construct which has withstood the scrutiny of several decades of empirical research. A commonly accepted definition of situation awareness has been proposed by Endsley (Endsley M. , 1995, p. 36): „Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future.“ Endsley further specifies situation awareness to occur at three consecutive stages, which are contingent upon the successful interplay of three key cognitive components: attention, working memory (WM) and long-term memory (LTM) (s. Figure 1).

The first step in achieving SA is the perception of status, attributes, and dynamics of goal-relevant stimuli in the environment. These are only perceived a) if the physical characteristics of these stimuli capture the human's attention (bottom-up processing) and/or b) if they are focused upon and determined to be relevant as indicated by operative goals that are stored in the WM and LTM (top-down processing).

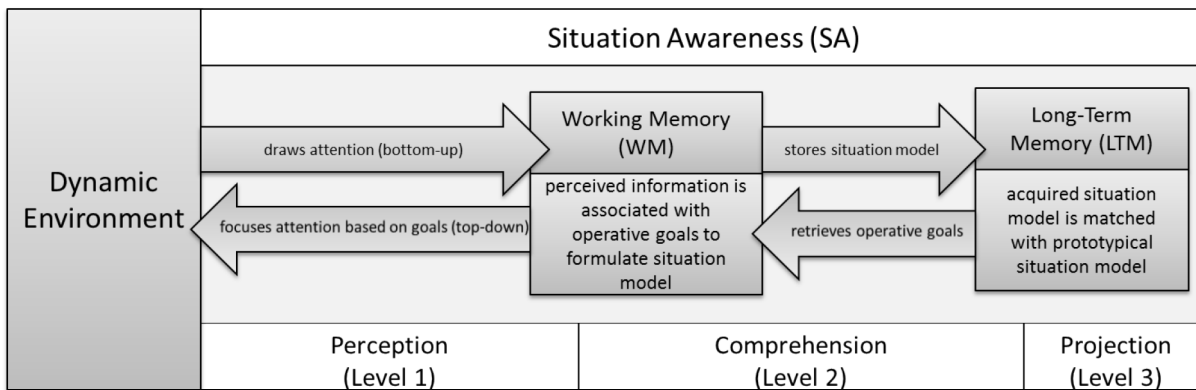


Figure 6. Schematic depiction of the cognitive processes involved in attaining situation awareness. Adapted from (Endsley M. , 1995).

In a second step towards achieving SA, perceived stimuli are synthesized and associated with operative goals in the WM, which have been retrieved from a larger number of goals stored in LTM. Hereby, the WM allows one to modify attention deployment on the basis of other perceived information or a change in active goals (Braune & Trollip, 1982). The goal-dependent synthesis of information stored in the WM leads to a comprehension of the current situation and the stimuli's relative significance in terms of achieving the goals.

In the third and final step towards achieving SA, future states of the elements in the environment may be projected based on the comprehension of the situation, leading to a timely and effective decision. For this purpose, a situation model which was created in the previous steps is stored in LTM and matched to schemata in memory that depict prototypical situations or states of the system model, which may be linked with goals that dictate further decision-making and actions. Hence SA equips the human agent with the ability to react appropriately in dynamic systems. In fact, numerous studies found a link between SA and effective performance (e.g. Gaba, Howard, & Small, 1995; Ma & Kaber, 2007; Gugerty & Tirre, 1996; Golightly, Wilson, Lowe, & Sharples, 2010).

Project SAAROS: Situation Awareness in Autonomous Service Robots

The aim of the proposed research program is to make a substantial theoretical and practical contribution towards improving the decision-making flexibility and hence adaptability of autonomously operating service robots using a distinctly interdisciplinary approach. To this end, a three-pronged approach is taken. Firstly, empirical data are gathered that reflect human processing of sensory input in the development of situation awareness, a cognitive process that was identified as pivotal to achieving adaptive behavior in dynamic situations. Secondly, a computational cognitive model which reflects these mechanisms is developed, using ACT-R, a cognitive architecture which was identified as featuring all necessary properties for the simulation of situation awareness (empirically supported cognitive plausibility,

appropriate learning mechanisms). Thirdly, the cognitive model is implemented in an autonomously operating NAO Next Gen robot (by Aldebaran) and evaluated for several scenarios that are relevant to service robotics and consequently adapted in order to reflect human-like adaptive behavior in dynamic environments. In the process of validating the developed model, systemic as well as structural influences which impact situation awareness in robotic systems are to be empirically uncovered and systematically documented. The following section details the working program of the Project SAAROS.

Step1: Identification of testing parameters

First, two benchmark scenarios and appropriate experimental setups within the domain of service robotics are defined, which serve to validate a computational SA model with experimental data collected from human participants and which may be juxtaposed with data provided by a cognitively equipped NAO robot in later stages of the project. Specifically, the first scenario is required for the development and validation of the SA model in a particular task domain of service robots (Step 6). A second scenario is devised at this stage, which will serve to evaluate the extent to which the developed model generalizes to other task domains within the field of service robotics (Step 7). For both scenarios, performance goals are specified that determine the aimed level of functionality for the robot once it is equipped with a validated cognitive model of SA.

Step2: Formulation of an ACT-R model of SA

A cognitively plausible ACT-R model of SA is formulated for the chosen context. In order to formulate an SA model within a cognitive architecture, the investigated scenario first needs to be deconstructed to its lowest operative level. For this purpose, the scenario is analyzed using goal-directed task analysis (GDTA), as proposed by Endsley (Endsley M. R., 1993). GDTA is a specific form of cognitive task analysis that focuses on identifying the goals and critical information needs for a particular task context. In essence, the product of this analysis is a hierarchy that specifically outlines the processes in which basic data needed by the user are integrated into higher SA levels of comprehension and projection. Second, based on the task analysis, an ACT-R model of SA is formulated.

Step 3: Acquisition of experimental data & model validation for human data

Following the GDTA (Step2), the model validation will occur in two stages, which are iteratively repeated until the model is considered to reflect human cognitive processes involved in SA satisfactorily: In a first step, the model is tested with regard to achievement and disruption of SA in humans. This includes experimental analysis with an opportunity sample of naïve participants. To measure the different levels of SA, a combination of tools will be used. Participants' visual focus of attention can be tracked using eye-tracking and motion tracking. The SA levels of comprehension and projection are assessed with two measurement techniques that have proven reliable and valid measures in numerous applications: SAGAT (Situa-

tion Awareness Global Assessment Technique) and SPAM (Situation Present Assessment Measure). With SAGAT (Endley, 1995), participants are intermittently queried, in the middle of a dynamic simulation, about the values of various parameters, whilst they are deprived of any feedback of the scene, so that the participant must rely on working memory to answer the questions. SPAM (Durso, Rawson, & Giroto, 2007) assesses the speed of accessing information while the scenario continues, so that the participant could seek the needed information rather than relying on memory. The SA measure, in this case is the time to respond. Statistical analyses will be conducted with quantitative measures in order to ascertain their degree of generalizability.

The formulated predictions are also tested with the ACT-R model. Functionally equivalent measures to those used with humans (eye-tracking, SPAM, SAGAT) can be applied to the measurement of SA in the ACT-R model: Visual focus of attention can be tracked and recorded during the programme's run-time. In addition, response times and various task performance indices and buffer content (i.e. working memory content) can be assessed. In addition, behavioural trends can be discerned with the ACT-R simulations. Upon conclusion of the analysis of the model's results, the ACT-R mechanisms may need to be adjusted and the hypotheses redefined. Hence, the testing process continues with stage (a) until the model achieves satisfactory performance (defined as overall model fit in performance measures of $R=.7$ or higher and comparable behavioural tendencies).

Step 4: Development of NAO/ACT-R Interface

Development work needs to be invested in order to equip an autonomously operating NAO robot, with an ACT-R model of SA. The higher level controls of NAO are realized by an embedded PC board in the robot's head, which runs on a LINUX operating system. Aldebaran provides an SDK-Toolkit called NAOqi which provides an interface to the hardware on low and high level bases, supporting different programming languages, including C/C++ and Python. ACT-R, on the other hand, is composed of a set of functions and algorithms implemented in Common Lisp, which is not supported by NAOqi. Hence, an interface will be devised for the robot, which enables the communication of the robot's controls with the perceptual-motor modules provided by ACT-R, for example, with use of CL Python, which is an open-source implementation of Python written in Common Lisp.

Step 5: Operationalisation of symbolic language with operative terms

With respect to the examined scenario and the results of the GDTA, operative terms need to be defined which translate the symbolic output of ACT-R into appropriate actions of the NAO robot and vice versa. This work package hence comprises two stages: Based on the GDTA (Step 2), a database is established which contains detailed instructions for the robot for each symbolism specified in the ACT-R SA model. For example, if ACT-R determines during its run-time that the robot should "scan for next visual cue", a corresponding translation is defined for

the robot's control and sensory processors, e.g. specifying the degrees of rotation of the robot's head, the area in which it scans and a choice of target stimuli. The second step involves the implementation of an interface linking the robot's API with an external PC on which the database is implemented. The decision to out-source this database was made to maintain fast run-times and at the same time ensure cognitive plausibility by considering these as subconscious processes, as humans do not, for example, consciously deliberate the degree to which they rotate their heads when they want to focus on an object of interest.

Step 6: Validation of ACT-R model in the NAO robot

Once the ACT-R model of SA developed in Step 2 has been validated with regard to human data and the model's symbolic language has been translated into operative terms for the robot, the model can be validated in the NAO robot. For this purpose, the hypotheses that were established during the model validation process (Step 3) are investigated with the NAO robot using comparable measures of SA to those described in Step 3. The model is considered valid for the presented scenario if it achieves satisfactory performance (defined as overall model fit in performance measures of $R=.7$ or higher and the display of comparable behavioural tendencies). If the model does not display satisfactory validity, possible causes are identified and the model and/or physical setup may be adjusted accordingly.

Step 7: Assessment of the model's ability to generalize to other task domains

In a final step, the extent to which the implemented model is task-specific is evaluated. For this purpose, the embodied ACT-R model is tested in a different test setup previously defined in Step 1, that requires the same basic skill set, but different reactions than the first scenario. From this final evaluation, systemic parameters may be delineated that can be identified as affecting SA in autonomously operating robots.

Conclusion

The multi-disciplinary project SAAROS at the Universität der Bundeswehr München can make contributions to three different domains: the empirical research on situation awareness in humans, the computational modeling of situation awareness using a cognitive architecture, and the embodiment of cognitive architectures in a robotic platform. Further investigations are planned to systematically assess the extent to which human cognitive processes may be applied to autonomously operating robots in order to produce adaptive behaviour in unforeseen situations.

References

International Federation of Robotics. (2012). Retrieved 10 12, 2012, from <http://www.ifr.org/service-robots/>

- Anderson, J. R., & Lebiere, C. (1998). *The Atomic Components of Thought*. Lawrence Erlbaum Associates.
- Bagnell, J., Bradley, D., Silver, D., Sofman, B., & Stentz, A. (2010). Learning for autonomous navigation. *IEEE Robotics & Automation Magazine*, 17(2), pp. 74-84.
- Baumann, M. R., & Krems, J. F. (2009). A comprehension based cognitive model of situation awareness. *Digital Human Modeling*, (pp. 192-201).
- Benjamin, D., Monaco, J. V., Lin, Y., Funk, C., & Lyons, D. (2012). Using a virtual world for robot planning. *SPIE Defense, Security, and Sensing*, (p. doi 10.1117/12.923446).
- Bratko, I. (2010). Comparison of machine learning for autonomous robot discovery. *Advances in Machine Learning I*, 262/2010, pp. 441-456.
- Braune, R., & Trollip, S. (1982). Towards an internal model in pilot training. *Aviation, Space and Environmental Medicine*(53), pp. 996-999.
- Decker, M., Dillmann, R., Dreier, T., Fischer, M., Gutmann, M., Ott, I., & Döhmman, I. (2011). Service robotics: do you know your new companion? Framing an interdisciplinary technology assessment. *Poiesis & Praxis. International Journal of Ethics of Science and Technology Assessment*, 8, pp. 25-44.
- Durso, F. T., & Sethumadhavan, A. (2008). Situation awareness: understanding dynamic environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3), pp. 442-448.
- Durso, F., Rawson, K., & Giroto, S. (2007). Comprehension and situation awareness. In F. Durso, R. Nickerson, S. Dumais, S. Lewandowsky, & T. Perfect, *Handbook of applied cognition* (2. ed., pp. 163-194). New York: Wiley.
- Endsley, M. (1995). Measurement of situation awareness in dynamic systems. *Human Factors*, 37(1), pp. 65-84.
- Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), pp. 65-84.
- Endsley, M. R. (1993). Situation awareness and workload: flip sides of the same coin. *Proceedings of the Seventh International Symposium on Aviation Psychology*, pp. 906-911.
- Gaba, D. M., Howard, S. K., & Small, S. D. (1995). Situation awareness in anesthesiology. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1), pp. 20-31.
- Golightly, D., Wilson, J. R., Lowe, E., & Sharples, S. (2010). The role of situation awareness for understanding signalling and control in rail operations. *Theoretical Issues in Ergonomics Science*, 1, pp. 84-98.

- Gonzales, C., & Dutt, V. (2007). Learning to control a dynamic task: a system dynamics cognitive model of the slope effect. *Proceedings of the 8th International Conference on Cognitive Modeling ICCM*, (pp. 61-66).
- Gonzalez, C., & Dutt, V. (2010). Instance-based learning models of training. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, (pp. 2319-2323).
- Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, pp. 591-635.
- Gugerty, L. J., & Tirre, W. C. (1996). Situation awareness: a validation study and investigation of individual differences. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, (pp. 564-568).
- Ma, R., & Kaber, D. B. (2007). Situation awareness and driving performance in a simulated navigation task. *Ergonomics*, 50(8), pp. 1351-1364. Retrieved November 21st, 2011, from <http://www.kivasystems.com>
- Peters, J., Morimoto, J., & Tedrake, R. (2009). Robot Learning. *IEEE Robotics & Automation Magazine*, 16(3), pp. 19-20.
- Sofge, D., Trafton, J. G., Cassimatis, N., Perzanowski, D., Bugajska, M., Adams, W., & Schultz, A. (2004). Human-robot collaboration and cognition with an autonomous mobile robot. *Proceedings of the 8th Conference on Intelligent Autonomous Systems (IAS-8)* (pp. 80-87). IOS Press.

Anthropomorphismus in der Gestaltung der Mensch-Roboter-Interaktion in industriellen Umgebungen

Sinem Kuz, Marcel Ph. Mayer und Christopher M. Schlick¹

Schlüsselwörter: Anthropomorphismus, Mensch-Roboter-Interaktion, Akzeptanz

Zusammenfassung

Unter Anthropomorphismus wird die Übertragung menschlicher Eigenschaften wie Gestalt oder Verhalten auf Nichtmenschliches verstanden. Die Anwendung auf einen Roboter soll die Arbeitssicherheit, Benutzerakzeptanz und Vertrautheit erhöhen und somit die Effektivität von Mensch-Roboter-Interaktionsprozessen optimieren. Im Umfeld der industriellen Robotik kann Anthropomorphismus speziell als Grundlage hinsichtlich der Planung von Bewegungen herangezogen werden. Deswegen wurde in einer experimentellen Untersuchung mittels einer virtuellen Umgebung das Antizipationsverhalten des Menschen auf anthropomorph modellierte Geschwindigkeitsprofile eines Industrieroboters untersucht. Die Ergebnisse zeigen, dass das anthropomorphe Profil zu deutlich kürzeren Prädiktionszeiten ohne eine Steigerung der Prädiktionsfehler führt.

Einleitung

Produzierende Unternehmen sind durch die steigende Variantenvielfalt von Produkten und immer kleiner werdende Losgrößen einem enormen Kosten-, Qualitäts- und Zeitdruck ausgesetzt. Es müssen anpassungsfähige Produktionssysteme entwickelt werden, die in unterschiedlichen Situationen flexibel eingesetzt werden können. Der Ansatz der Hochautomatisierung stößt dabei häufig an seine Grenzen. Um die notwendige Flexibilität zu erreichen sollte der Mensch mit seinen kognitiven Fähigkeiten in den Produktionsprozess aktiv eingebunden werden. Somit können die spezifischen Fähigkeiten und Fertigkeiten von Mensch und Maschine optimal kombiniert werden. Neben der essentiellen Fragestellung nach adäquaten Sicherheitskonzepten, welche den Schutz der Arbeitsperson gewährleisten, sind für eine effektive Zusammenarbeit die Betrachtung von Einflussfaktoren auf die Interaktionsbereitschaft des Menschen und die Akzeptanz von roboterbasierten Systemen als Interaktionspartner ebenfalls zentrale Fragestellungen. In diesem Kontext wird häufig der Begriff „Anthropomorphismus“ verwendet, welcher die Übertragung von menschlichen Eigenschaften oder menschlichen Verhaltens auf Nichtmenschliches beschreibt. Menschen neigen generell zu dieser Methode der „Vermenschlichung“, da sie versuchen auf diese Weise das Verhalten ihres nicht-menschlichen Gegenübers zu interpretieren (Duffy, 2003). Durch die Übertragung von menschähnlichen Merkmalen auf technische Systeme wie Roboter soll somit eine höhere

¹ Lehrstuhl und Institut für Arbeitswissenschaft der RWTH Aachen

Benutzerakzeptanz und Vertrautheit erzielt werden, wodurch die Effektivität und Sicherheit von Mensch und Roboter in Interaktionsszenarien weiter verbessert werden soll (Oleson, 2011). Hierzu wird derzeit eine Studie durchgeführt, in der auf den Einfluss des Geschwindigkeitsprofils des Endeffektors sowie der kinematischen Elemente fokussiert wird. Die ersten Ergebnisse dieser Studie werden im Rahmen dieses Beitrags vorgestellt.

Anthropomorphismus und technische Systeme

Luczak et al. (2003) haben bereits früh die Auswirkungen von Anthropomorphismus im Zusammenhang mit technischen Geräten erforscht und konnten nachweisen, dass Menschen solche Geräte mit anthropomorphen Eigenschaften in der Regel als Helfer oder Freunde wahrnehmen und diese freundlicher als einfache Geräte behandeln. Hinds et al. (2004) haben sich auf Roboter fokussiert und untersuchten, wie sich ein anthropomorphes Aussehen auf die menschliche Wahrnehmung während der Ausführung einer gemeinsamen Aufgabe mit dem Roboter auswirkt. Demnach stellte sich heraus, dass Menschen in Situationen, in denen es um die Übertragung von Verantwortung geht, Roboter mit anthropomorphen Eigenschaften als geeigneter empfinden. In einer weiteren Studie untersuchten Krach et al. (2008) die Auswirkungen von Menschähnlichkeit auf die wahrgenommene Intelligenz in der Interaktion mit Computersystemen. Das Basis-System innerhalb der Studie war ein einfaches Computerspiel. Anthropomorphismus variierte dabei von einem einfachen Laptop bis hin zu einem menschlichen Gegenüber. Ergebnisse zeigten eine positive Korrelation zwischen dem Grad der Menschenähnlichkeit und der wahrgenommenen Intelligenz. Nichtsdestotrotz ist zu berücksichtigen, dass eine höhere Ähnlichkeit zum Menschen zwar zu einer gesteigerten Vertrautheit führt, gleichzeitig aber auch negative Konsequenzen nach sich ziehen kann. Die Theorie des „Uncanny Valley“ von Mori (1970) beschreibt hierzu zunächst einen Anstieg der Vertrautheit mit zunehmender Menschähnlichkeit, wobei ab einem gewissen Punkt bevor der Grad an Menschähnlichkeit das Maximum erreicht die Vertrautheit abnimmt und in Ablehnung umschlägt. Allerdings gibt es bisher keinen wissenschaftlich fundierten Nachweis der von diesem Modell postulierten Effekte. (Geller, 2008).

Da traditionelle Ansätze in der industriellen Robotik bisher rein effiziente und stark automatisierte Lösungen fokussierten, wurde der Ansatz des Anthropomorphismus in Mensch-Roboter-Arbeitssystemen kaum angewendet. Eine Erhöhung der Menschähnlichkeit würde aber gemäß dem Uncanny Valley-Modell zunächst einen positiven Effekt auf die Vertrautheit haben. Die Arbeiten von Zanchettin (2012) konzentrieren sich hierzu auf die Bewegungssteuerung eines Industrieroboters während Manipulationsaufgaben, wodurch für jede beliebige Position des Endeffektors eine anthropomorphe Stellung des Ellbogengelenks bestimmt werden kann. Hierbei bleiben die menschlichen Bewegungsbahnen bzw. Trajektorien jedoch unbeachtet. Bei Anwendung der anthropomorphen Stellungen auf einen Sieben-Achs-Dualarmroboter konnte in empirischen Studien eine Minderung des Be-

anspruchsniveaus der Teilnehmer nachgewiesen werden, die mit diesem Dualarmroboter zusammenarbeiteten. HUBER et al. (2008) untersuchten wie sich menschliches Kooperationsverhalten durch modifizierte Robotertrajektorien auf die Mensch-Roboter-Zusammenarbeit übertragen lässt. Als Basis ihrer Untersuchungen verwendeten sie einen Übergabeprozess, bei dem sie die Mensch-Mensch-, Mensch-Knickarmroboter- und Mensch-Humanoid-Kooperation verglichen. Gleichzeitig nutzten sie eine an menschlichen Bewegungen orientierte Bahn, die sie nach der Maxime eines möglichst ruckfreien Ablaufs modellierten und verglichen diese mit einer Trajektorie, die sich aus einem trapezförmigen Geschwindigkeitsprofil ergab. Die Auswertung zeigte, dass die Reaktionszeiten, d.h. die Dauer zwischen dem Moment des Übergebens und dem des Übernehmens, bei der menschlich gestalteten Bahn geringer waren als im trapezförmigen Fall. Ein ähnliches Ergebnis ergab der Vergleich vom industriellen und humanoiden Roboterdesign, da auch in diesem Fall das anthropomorphe Erscheinungsbild zu einer verkürzten Reaktionszeit führte. Hier basieren die Ergebnisse auf berechneten Trajektorien und nicht den natürlichen anthropomorphen Bewegungsbahnen oder Geschwindigkeitsprofilen. Diesbezüglich konnte in einer ersten Studie von Kuz et al. (2013) mit einem Industrieroboter in einer virtuellen Umgebung nachgewiesen werden, dass anthropomorph modellierte Bewegungsbahnverläufe im Vergleich zu konventionellen Trajektorien eines industriellen Knickarmroboters die menschliche Vorhersagegenauigkeit von Zielpositionen positiv beeinflussen und die Reaktionszeit signifikant verkürzen kann. Ein Vergleich der Geschwindigkeitsprofile blieb jedoch unbeachtet und soll deswegen im Rahmen dieser Studie näher untersucht werden.

Methodik

Die Bewegungen des Industrieroboters in der virtuellen Umgebung im Rahmen dieser Studie unterscheiden sich im Gegensatz zur Vorstudie von Kuz et al. (2013) nur hinsichtlich des Geschwindigkeitsprofils und nicht des Bahnverlaufs. Hierbei kam ein Modell eines KUKA KR 30 Jet Protalroboters mit über-Kopf-Aufhängung zum Einsatz (siehe Abb.1). Es wurden zunächst mittels eines Infrarot-Tracking-Systems menschliche Bewegungsbahnen während Platzierungsbewegungen erfasst und für die Übertragung auf die virtuelle Versuchsumgebung angepasst. Für die zweite Bewegungsart wurde der menschliche Bahnverlauf angepasst und mit einem konstanten Geschwindigkeitsprofil kombiniert. Während des Versuchs wurden die Teilnehmer gebeten den Knickarmroboter in der virtuellen Umgebung zu beobachten und den Ablageort der präsentierten Bewegung möglichst schnell vorherzusagen. Hierzu sollte der Proband mittels Maus-Eingabe das Feld auf dem virtuellen Schachbrett auswählen (siehe Abb.1), von dem er glaubt, dass dies die Zielposition der Bewegung sei. Jedes der Felder wurde sowohl durch menschenähnliche als auch konstante Geschwindigkeiten angefahren, wobei die Reihenfolge der Bewegungsabläufe für jeden Teilnehmer permutiert wurde. Nach jeder Eingabe des Probanden wurde die virtuelle Szene unterbrochen, so dass keine Rückmel-

dung über den Wahrheitsgehalt der Prädiktion erfolgte. Die Zeit zwischen Beginn der Bewegung und der Eingabe des Probanden wurde softwareseitig erfasst und ausgewertet.

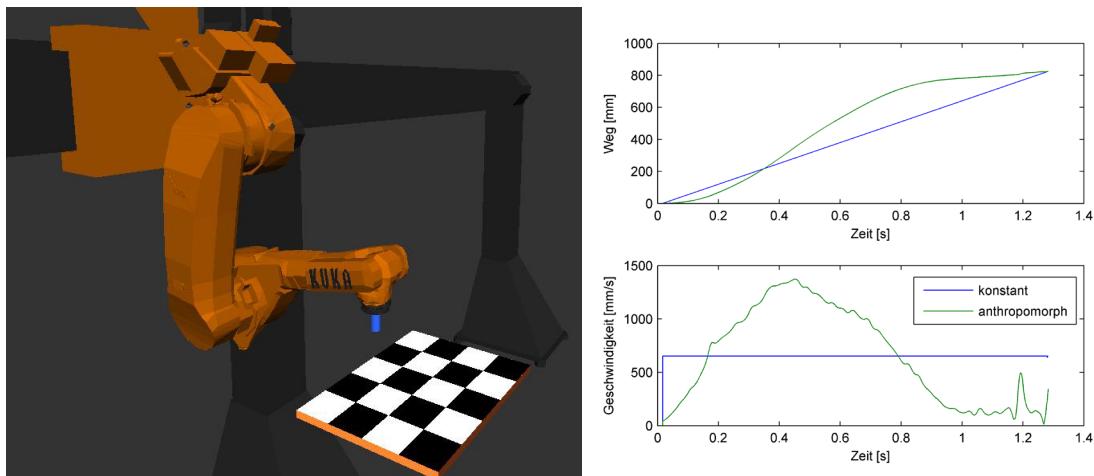


Abb. 1: Darstellung der im Experiment verwendeten Versuchsumgebung (links) sowie der Weg- bzw. Geschwindigkeitsprofile (rechts)

An der Studie nahmen bisher insgesamt 22 männliche Probanden mit technischer Ausbildung im Alter von 23 bis 37 Jahren (MW: 27,18 Jahre, SD: 3,16 Jahre) teil.

Als unabhängige Variablen wurden die Robotertrajektorien mit zwei Geschwindigkeitsprofilen (siehe Abb. 1 rechts) und die Feldposition mit 20 Stufen betrachtet. Als abhängige Variablen wurden schließlich die Prädiktionsdauer und die Anzahl der Prädiktionsfehler untersucht.

Ergebnisse

Die statistische Auswertung wurde mit einer ANOVA mit meßwiederholung auf einem Signifikanzniveau von $\alpha=0,05$ durchgeführt. Die ermittelten Ergebnisse zeigen, dass sowohl die Bewegungsart ($F(1, 21) = 95,213$, $p < 0,001$) als auch die Position ($F(7,866;165,177) = 52,746$, $p < 0,001$) einen signifikanten Einfluss auf die Prädiktionszeit haben. Beide Haupteffekte können aufgrund der ordinalen Interaktion ($F(5,096; 107,021) = 5,725$, $p < 0,001$) getrennt voneinander interpretiert werden. Der Effekt bzgl. der Felder auf dem Schachbrett ist erwartungskonform, da die Bewegungen zu weiter entfernten Feldern tatsächlich mehr Zeit bei der Vorhersage in Anspruch nehmen. Bezüglich der Bewegungsart zeigt sich, dass das anthropomorphe Geschwindigkeitsprofil zu einer signifikant schnelleren Prädiktion führt. So liegt die durchschnittliche Reaktionszeit beim konstanten Geschwindigkeitsprofil bei 2,02 s, wohingegen sie bezüglich des anthropomorphen Profils bei 1,68 s liegt (siehe Abb. 2).

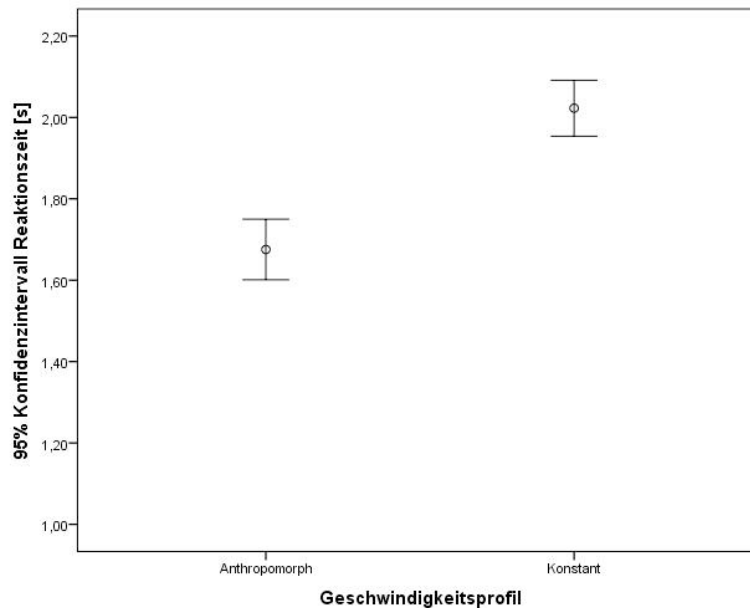


Abb. 2: Die erhobenen Prädiktionszeiten

Um einen Speed-Accuracy Tradeoff auszuschließen, wurden zusätzlich die Prädiktionsfehler ausgewertet. Der Unterschied zwischen den beiden Geschwindigkeitsprofilen ist nicht signifikant ($T(21)=1,501$, $p>0,05$), so dass kürzere Reaktionszeiten nicht zu mehr Fehlprädiktionen führen.

Ausblick

Gegenstand dieses Beitrags ist eine experimentelle Studie, in der die Auswirkungen von menschähnlichen Bewegungsbahnen eines Portalroboters auf die Vorhersagegenauigkeit von Zielpositionen untersucht wurden. Die Ergebnisse zeigen, dass das anthropomorphe Profil zu einer signifikant kürzeren Prädiktionszeit führt. Dabei führte die schnellere Bearbeitungszeit nicht zu mehr Prädiktionsfehlern, so dass ein Speed-Accuracy Tradeoff zumindest statistisch ausgeschlossen werden kann. Auf Basis der ermittelten Ergebnisse wird eine weitere Studie vorbereitet, in der die Auswirkungen von anthropomorph modellierten Bewegungsbahnen eines realen Montageroboters auf die Vorhersagbarkeit des Menschen untersucht werden sollen. Dabei liegt der Fokus erneut auf den Vergleich zwischen den anthropomorphen und den konventionellen Trajektorien während Platzierungsbewegungen.

Literatur

- Duffy, B. (2003): Anthropomorphism and The Social Robot. Special Issue on Socially Interactive Robots, Robotics and Autonomous Systems, 42 (3-4), 177–190.
- Geller, T. (2008). *Overcoming the Uncanny Valley*. Computer Graphics and Applications (11-17). IEEE

- Hinds, Pamela J., Roberts, Teresa L., and Jones, H. (2004). *Whose job is it anyway? a study of human-robot interaction in a collaborative task*. *Hum.-Comput. Interact.* 19, 1 (151-181). doi: 10.1207/s15327051hci1901&2_7
- Huber, M, Rickert, M., Knoll, A., Brandt, T., Glausauer, S. (2008). *Human-robot interaction in handing-over tasks*. In *Proc. of RO-MAN*. München (107-112): IEEE.
- Krach, S., Hegel F., Wrede B., Sagerer G., Binkofski F., Kircher T. (2008). Interaction and perspective taking with robots investigated via fMRI. *PLoS ONE*. 3(7): e2597. doi: 10.1371/journal.pone.0002597
- Kuz, S., Mayer, M. Ph., Heinicke, A. und Schlick, C.M. (2013). Auswirkung anthropomorph modellierter Trajektorien eines Montageroboters auf die Prädiktion der Endeffektorposition. In *Chancen durch Arbeits-, Produkt- und Systemgestaltung – Zukunftsfähigkeit für Produktions- und Dienstleistungsunternehmen*, Bericht zum 59. Kongress der Gesellschaft für Arbeitswissenschaft. Dortmund (433-436): Gesellschaft für Arbeitswissenschaft e.V., GfA-Press
- Luczak, H., Rötting, M., Schmidt, L.(2003). Let's talk: anthropomorphization as means to cope with stress of interacting with technical devices. In *Ergonomics*, London (1361-1374).
- Mori, M. (1970). *The Uncanny Valley*. In *Energy*, 7(4), S. 33-35, Translated by Karl F. MacDorman and Takashi Minato
- Oleson, K.E., Billings, D.R., Kocsis, V., Chen, J.Y.C., & Hancock, P.A. (2011). Antecedents of trust in human-robot collaborations. In *Proceedings of the International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*. Orlando, FL, USA (175-178): IEEE
- Zanchettin, A. M. (2012). Human-centric behavior of redundant manipulators under kinematic control, PhD Thesis. Politecnico di Milano.

Kooperative Automation im Labor - Vergleich von manueller und teilautomatisierter Prozessüberwachung und -führung

Charlotte von Bernstorff¹ und Jens Nachtwei²

Funktionsallokation, Operateur, Leistung, Formen kooperativer Automation

Zusammenfassung

Kooperative Automation ist die heute in der Human Factors Forschung favorisierte Automationsstrategie, bei der sich Operateur und Automation die Verantwortung für Funktionen gegenseitig dynamisch übergeben können. Um Kriterien für eine solche Zusammenarbeit zu identifizieren, z.B. wer soll wann die Kontrolle haben, ist es sinnvoll, den Einfluss von Automationsstufen auf die Leistung iterativ zu untersuchen. In dieser Studie wurde die manuelle mit einer teilautomatisierten Prozessüberwachung und -führung in Bezug auf die Leistung in einer spezifischen Simulationsumgebung verglichen: Zwei Personen, sogenannte Navigatoren, führen eine kooperative Trackingaufgabe aus und werden dabei von einer dritten Person, dem Operateur, (manuell oder teilautomatisiert) unterstützt. In der teilautomatisierten Bedingung produzierten die Navigatoren weniger Fehler, als in der manuellen Bedingung und sind zudem weniger angestrengt. Ein Moderatoreffekt aus Bedingung mal Anstrengung auf die Leistung ergab, dass der positive Zusammenhang von Anstrengung und Leistung in der teilautomatisierten Bedingung abschwächt wird. Das Vorgehen erlaubt valide Rückschlüsse für Kriterien dynamischer Funktionsallokation.

Operateur und Automatik als Kooperationspartner – aber wie?

Der Trend zur Automatisierung von Handlungen und Informationsverarbeitungsprozessen in vielfältigen Domänen (Parasuraman & Wickens, 2008) fordert Antworten auf die Frage, welche Funktionen/Aufgaben auf welche Weise an Mensch und Automation vergeben werden sollten – die Frage nach der Funktionsallokation (FA). In klassischen Ansätzen zur Realisierung der FA (z.B. Fitts, 1951) wurden Funktionen bereits als Designentscheidung unwiderruflich an entweder Automation oder Operateur vergeben. Damit wird die Interaktion zwischen Operateur und Automation auf eine trennscharfe Aufgabenteilung „Du tust dies, ich tue das“ (Dekker und Woods, 2002, S. 7) reduziert. Solche Ansätze gelten heute als zu starr und deswegen veraltet. Stattdessen wird eine FA *während* des Interaktionsprozesses (Sheridan & Parasuraman, 2006) gefordert. Bei dieser sogenannten dynamischen FA (Kaber, Riley, Tan & Endsley, 2001) oder Kooperativen Au-

¹ Technische Universität Berlin, Zentrum für Mensch Maschine Systeme (ZMMS)

² Humboldt-Universität zu Berlin, Institut für Psychologie

tomation (Flemisch et al., 2008), können sich Operateur und Automation die Verantwortung für den Prozess *während* der Prozessüberwachung und -führung teilen bzw. gegenseitig übergeben. Das Automationslevel kann sowohl adaptiv (durch die Automation) oder adaptierbar (durch den Operateur) verändert werden (Hildebrandt & Harrison, 2002) und sich daher besser an Bedürfnisse und Zustände des Operators anpassen. In der Literatur wird bisher jedoch unzureichend beschrieben, wie genau eine solche Zusammenarbeit zwischen Automation und Operateur aussehen kann, beispielsweise wer von beiden über den Wechsel des Automationslevels entscheidet (Sauer, Kao & Wastell, 2012) oder nach welchen Kriterien dieser Wechsel initiiert werden soll (Kaber et al., 2001; Sheridan & Parasuraman, 2006). Nach bereits existierenden Befunden (z.B. Lemoine, Debernard & Millot, 1995) bevorzugen Operateure es scheinbar, selbst über den Wechsel zu entscheiden, also die Kontrolle über die Automation behalten zu wollen. Wie sich Operateurs- vs. Automatik-induzierte Wechsel tatsächlich auf die Prozessführungsleistung auswirken, ist bisher nicht geklärt. Neben Hinweisen *für* die Kontrolle der Automatik durch den Operateur (Hancock, 2007; Sauer et al., 2012) führt diese nach einer aktuellen Studie von Sauer, Nickel & Wastell (2013) nicht unbedingt zur Leistungssteigerung. Die widersprüchlichen Ergebnisse sind sicher auch mit dem Wirken diverser Mediator- und Moderatorvariablen auf die Beziehung zwischen dynamischer Automation und Prozessführungsleistung zu erklären, z.B. auf Umgebungsbedingungen wie Lautstärke (Sauer et al., 2013) oder den Workload des Operators (Parasuraman & Wickens, 2008). Um sich der Zusammenarbeit zwischen Operateur und Automation Schritt für Schritt zu nähern, wurde in einem aktuellen Forschungsprojekt an der Humboldt-Universität zu Berlin daher zunächst der einfachste Fall einer Mensch-Automatik-Interaktion (Kienle et al., 2009) unter der Berücksichtigung möglicher Moderatorvariablen untersucht: das Einschalten und Ausschalten von (nicht adaptiven) Automatikfunktionen durch den Operateur. Dafür wurde eine spezielle Simulationsumgebung um automatische Funktionen erweitert, in welcher ein Operateur einen komplexen dynamischen Prozess – simuliert durch ein kooperatives Tracking – optimiert. Diese wird im folgenden Abschnitt beschrieben.

Manuelle vs. teilautomatisierte Prozessüberwachung und -führung im Labor

Im Rahmen einer an der Humboldt-Universität zu Berlin entwickelten Simulationsumgebung erhalten Probanden in der Rolle von Operateuren die Aufgabe, einen komplexen, dynamischen Prozesses zu überwachen und zu steuern (Abbildung 1). Dieser komplexe, dynamische Prozess wird simuliert durch eine kooperative Tracking-Aufgabe zweier weiterer Probanden: Zwei sogenannte Navigatoren steuern gemeinsam ein Fahrobjekt eine virtuelle Fahrstrecke entlang. Sie nutzen dafür jeweils einen Joystick mit 50% Steueranteil. Navigator 1 soll vor allem schnell, dabei genau, Navigator 2 soll vor allem genau, dabei schnell, fahren. Die Navigatoren gehen jedoch fälschlicherweise davon aus, identisch instruiert worden zu sein.

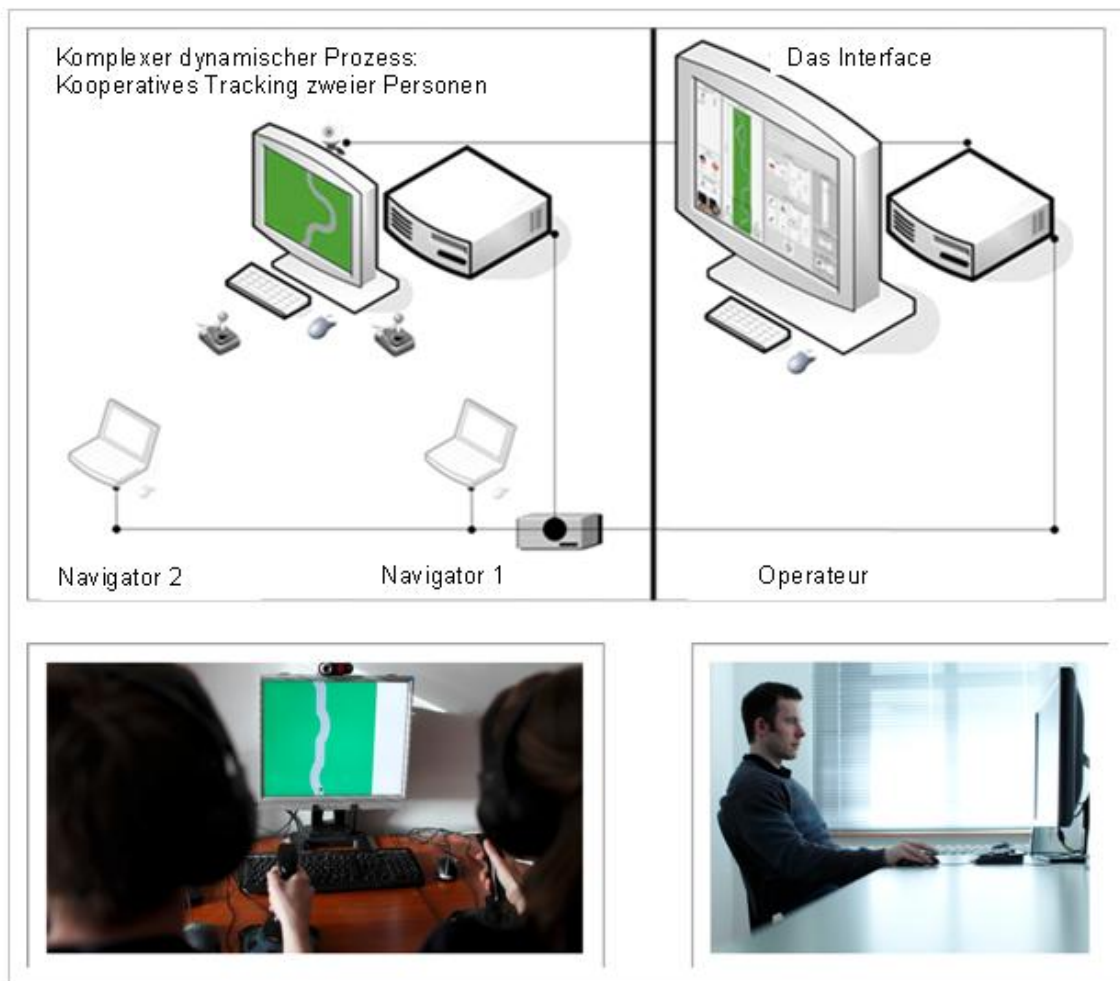


Abb. 1: Aufbau der Simulationsumgebung. Links: Zwei Navigatoren steuern ein Fahr-objekt mithilfe von Joysticks. Jeder hat einen Steueranteil von 50%. Fahrleistung, Anstrengung und Videobild beider Personen werden dem Operateur zurück gemeldet. Rechts: Der Operateur nutzt ein Interface, um beide Personen zu beobachten und deren Fahrleistung (Geschwindigkeit und Genauigkeit) zu optimieren. Die schwarze Linie im Bild oben repräsentiert die räumliche Trennung.

Insgesamt absolvieren die Navigatoren 11 Fahrten, davon 8 individuelle und kooperative Trainingsdurchläufe von ansteigender Dauer und Schwierigkeit sowie schließlich 3 messrelevante Fahrten auf höchstem, jeweils vergleichbarem Schwierigkeitsniveau. Nach jeder Fahrt schätzen beide Navigatoren auf einer computerbasierten Skala, einer adaptierten Form der Rating Scale Mental Effort (RSE, Zijlstra & van Dorn, 1985) ein, wie anstrengend sie die jeweilige Fahrt empfunden haben. Weiterhin werden beide Personen darüber informiert, dass eine Software zur Optimierung ihrer Fahrleistung eingesetzt wird, welche visuelle und auditive Hinweise (über Kopfhörer abgespielte Audiodateien) zum Fahrverhalten an beiden Navigatoren rückmelden und direkte Eingriffe in das Fahrverhalten tätigen kann. Tatsächlich handelt es sich bei der "Software" jedoch um eine weitere Versuchsperson im Nebenraum, den erwähnten Operateur. Dieser soll in einem spezifischen Interface (Abbildung 2) Informationsquellen (z.B. Streckenvorschau, Echtzeitvideo und Anstrengungswerte der beiden Personen) und Eingriffsmöglichkeiten (z.B. visuelle

und auditive Geschwindigkeitshinweise, Reduktion der Fahrtgeschwindigkeit, Veränderung der Steueranteile (von 50%-50% z.B. zu 25%-75%) nutzen, um beide Personen zu beobachten und deren Fahrleistung zu optimieren.



Abb 2: Das Interface. Der Operateur kann den Fahrprozess sowie die beiden Personen selbst beobachten (linke Seite) und Hinweise (z.B: Warnungs-, Richtungs-, Geschwindigkeitshinweise oder Lob) oder direkte Eingriffe (z.B. Reduktion des Geschwindigkeitsl

In einer früheren Untersuchung konnte der Operateur alle Eingriffe ausschließlich manuell, also durch aktiven Linksklick auf eine Schaltfläche tätigen. Für die zugrunde liegende Studie wurde die Simulationsumgebung um teilautomatische Funktionen erweitert (Abbildung 3):

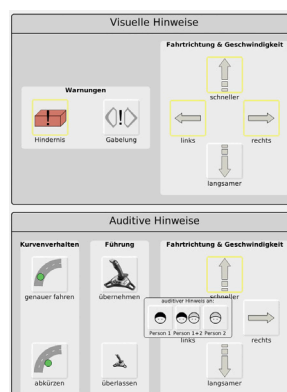


Abb 3: Visueller und auditiver Hinweis "Schneller fahren", Visuelle Hinweise "Links" und "Rechts fahren" werden von der Automatik ausgeführt (sichtbar durch Umrandung der jeweiligen Schaltflächen).

Die visuellen und auditiven Richtungs- und Geschwindigkeitshinweise sowie die visuellen Warnhinweise kann der Operateur entweder, wie bisher, manuell aktivieren, oder (durch Rechtsklick auf die jeweilige Schaltfläche) eine automatische Unterstützung aktivieren, die statt seiner den jeweiligen Hinweis ausführt. Die automatische Unterstützung übernimmt dann so lange diese Aufgabe, bis der Operateur sie (durch Links- oder Rechtsklick) wieder deaktiviert. Eine Rückmeldung, ob die automatische Unterstützung aktiv ist, erhält der Operateur durch eine (gelbe) Umrandung der jeweiligen Schaltfläche. Die Operateure entscheiden selbständig, ob und in welchen Situationen sie automatische Funktionen aktivieren. Es interessiert, inwiefern dieser einfachste Fall der kooperativen Automation (Ein- und Ausschalten der Automation) die Leistung des Gesamtsystems beeinflusst und insbesondere ob und wie sich die Leistung des Operators im Vergleich zur rein manuellen Bedingung verändert. Die Prozessüberwachungs- und -führungsleistung des Operators (Optimierung des kooperativen Trackings der beiden Navigatoren) wird dabei u.a. über die Genauigkeit des Trackings erfasst. Orientiert an Wiener (1949), der die Trackingqualität als erster so definierte, wird die Genauigkeit als mittlere Summe der Abweichungen des Fahrobjekts von der vorgegebenen Strecke definiert. Darüber hinaus interessieren diverse weitere leistungsrelevante Maße, z.B. die Anstrengung der Navigatoren oder der Workload des Operators. Die Leistungsmaße sowie alle Eingriffe des Operators bzw. der Automatik werden als Logfiles mit einer Auflösung von 39 ms gemessen und erlauben somit eine hoch reliable Erfassung der Leistung mit und ohne Automation. In diesem Artikel wird als Ergebnisauszug der Zusammenhang der Art Prozessüberwachung und -führung mit Fehlern im Tracking dargestellt.

Präzision durch Automation: Der Zusammenhang von automatisch gestützter Prozessüberwachung und -führung auf die Leistung

Insgesamt 81 Operateure (und entsprechend 162 Navigatoren, da jeder Operateur mit einem unabhängigen Team aus zwei Navigatoren interagiert) nahmen an der Studie teil, davon 40 Personen (16 weibliche und 14 männliche) in der schon zwei Jahre zuvor durchgeführten, rein manuellen und 41 Personen (21 weibliche und 20 männliche) in der teilautomatisierten Bedingung. Das Alter aller Probanden wurde kontrolliert und lag zwischen 18 und 35 Jahren. Der Bildungsabschluss war mindestens Abitur. Studenten der Psychologie, Linkshänder und Personen mit Rot-Grün-Schwäche wurden ausgeschlossen.

Um den Zusammenhang der Art der Prozessüberwachung und -führung (manuell vs. teilautomatisiert) mit der Leistung (hier: Fehler im Tracking) sowie den Effekt möglicher Moderatoren (z.B. Anstrengung der beiden Navigatoren) zu überprüfen, wurde die Multiple Lineare Regression (MLR) eingesetzt. Gegenüber der vielfach angewandten Varianzanalyse (ANOVA) wird dieses Verfahren insbesondere dann empfohlen, wenn es um die Behandlung von Mediator- und Moderatorvariablen geht (Cohen, Cohen, West & Aiken, 2003). Dieses Verfahren erlaubt die systemati-

sche Berechnung von (inkrementellen) Varianzanteilen einzelner Variablen, so auch von Interaktionstermen (Stade, Meyer, Niestroj & Nachtwei, 2011). Es wurde angenommen, dass Operateure in der teilautomatisierten Bedingung, die also mit einer Automatik zusammen arbeiten durften, die Trackingleistung der Navigatoren verbessern sollten. Im Vergleich zu der Bedingung, in welcher der Operateur ausschließlich manuell überwachen und eingreifen, somit auf keine automatische Unterstützung zurückgreifen konnte, sollte sich die Zusammenarbeit mit der Automatik in weniger Fehlern und weniger Zeit im Tracking äußern. Die Option, Aufgaben an eine Automatik abzugeben, sollte den Operateur entlasten und ihm ermöglichen, die Navigatoren und deren Zustand per Videobild und deren jeweils zuletzt angegebener Anstrengungswerte besser zu beobachten und zu unterstützen. Somit sollten auch die beiden Navigatoren selbst weniger angestrengt sein. Tatsächlich wirkt sich die automatische Unterstützung positiv auf die Genauigkeit des Trackings aus. Die Ergebnisse der Multiplen Regression sind in Tabelle 1 abgebildet.

Tab.1: Ergebnisse der Multiplen linearen Regression: Zusammenhang zwischen Prozessführung und Leistung (Fehler im Tracking)

		Regressions- koeffizient B	Standard- fehler	Beta	Teil- korr.	<i>p</i>
Modell 1	(Konstante)	4.43	0.10			
	Prozessführung	-1.52	0.21	-0.64	-.64	.00
Modell 2	(Konstante)	4.43	0.09			
	Prozessführung	-1.25	0.22	-0.53	-.47	.00
	Anstrengung	0.02	0.01	0.25	.22	.01
Modell 3	(Konstante)	4.27	0.11			
	Prozessführung	-1.36	0.21	-0.57	-.50	.00
	Anstrengung	0.01	0.01	0.13	.11	.17
	Aut*Anstr	-0.04	0.01	-0.27	-.25	.02

Prädiktor 1: Prozessführung (manuell vs. teilautomatisiert, aufsteigend kodiert)

Prädiktor 2: Anstrengung der Navigatoren in der vorherigen Fahrt

Prädiktor 3: Interaktion aus Prozessführung und Anstrengung

Kriterium: Leistung (Fehler im Tracking)

$R^2 = .41$ für Modell 1; $\Delta R = .05$ für Modell 2 ($p < .05$), $\Delta R = .06$ für Modell 3 ($p < .05$).

Die
Teil-

korrelationen (vorletzte Spalte der Tabelle) repräsentieren den Zusammenhang von Prozessführung (manuell vs. teilautomatisiert), Anstrengung (RSE-Werte) sowie der Interaktion aus Prozessführung (Aut*Anstr) mit der Leistung bzw. der Genauigkeit des Trackings (operationalisiert über die mittlere Summe der Abweichungen des Fahrobjekts von der Strecke). In der letzten Spalte *p* wird das jeweilige Signifikanzniveau angegeben. Die Navigatoren machen signifikant weniger Fehler im Vergleich zu der Bedingung, in welcher der Operateur auf sich allein gestellt ist und alle Eingriffe manuell tätigen muss. Das belegt die hohe Teilkorrelation von $r = -.50$ ($p < .001$) im Modell 3. Grundsätzlich werden mehr Fehler gemacht, je hö-

her die Anstrengung der beiden Navigatoren ist (Teilkorrelation der Anstrengung von $r = .22$, $p < .05$ im Modell 2). Dieser Effekt wird jedoch interessanter Weise aufgehoben, wenn die Interaktion aus Bedingung und Anstrengung als dritter Prädiktor in das Modell mit eingeht. Der resultierende Moderatoreffekt (Teilkorrelation des Interaktionsterms von $r = -.25$, $p < .05$) zeigt, dass der Zusammenhang zwischen Anstrengung und Leistung in der teilautomatisierten gegenüber der manuellen Bedingung abgeschwächt wird. Dies lässt sich mithilfe eines Mittelwertvergleichs der Anstrengung über beide Bedingungen erklären: Ein t-Test zeigt, dass die Navigatoren in der teilautomatisierten Bedingung im Mittel eine sehr viel geringere Anstrengung von nur 39,93% ($SD = 10,59$; Spannweite = 43,0) aufweisen und sich damit signifikant ($T = 4.478$, $p < .05$) von den Navigatoren in der manuellen Bedingung mit einer mittleren Anstrengung von 53,54 % ($SD = 15,92$; Spannweite = 76,25 (53 nach Ausreißer-Analyse) unterscheiden. Wenn die Prozessüberwachung und -führung rein manuell erfolgt, existiert ein starker Zusammenhang zwischen Anstrengung und Leistung, dagegen wird der Effekt von Anstrengung auf die Leistung durch den Einfluss der teilautomatischen Unterstützung abgepuffert.

Mehr Aufmerksamkeit durch Aufgabendelegation?

Aus der Vielzahl an subjektiven und objektiven Messergebnissen, die mit der Simulationsumgebung erzielt werden können (z.B. zum Einfluss von Persönlichkeitsmerkmalen, zur Art und Häufigkeit von Eingriffen und Interaktionen zwischen Operateur und Navigatoren), wurde ein zentraler Befund berichtet, der den positiven Einfluss einer teilautomatisierten Prozessüberwachung und -führung auf die Prozessleistung (Genauigkeit des Trackings) beschreibt. Die hohe Effektstärke von .50 zeigt, wie stark Operateure bei der Prozessüberwachung und -führung von der optionalen Aufgabendelegation an eine Automatik, im Vergleich zu der Bedingung, in der alle Aufgaben manuell zu bewältigen sind, profitieren. Dabei ist hervorzuheben, dass der Operateur und die ihn unterstützende Automatik im Gegensatz zu den Navigatoren kaum bzw. gar keinen direkten Einfluss auf den Fahrprozess haben. Es ist beachtlich, dass somit allein durch die indirekte Wirkung der teilautomatisierten Prozessüberwachung und -führung des Operateurs ein derart starker Effekt zu verzeichnen ist. Offenbar haben die Operateure in leistungsförderlicher Weise mit der Automatik zusammen gearbeitet. Das spricht nicht zuletzt für die Antizipationsleistung der Entwickler dieser automatischen Unterstützung (Cummings & Thornburg, 2011), auf deren Intelligenz das Funktionieren – und die adäquate Nutzung – einer Automatik zurückzuführen ist (Norman, 2007). Denn eine erfolgreiche Prozessüberwachung und -führung durch Operateure hängt immer auch von der Frage ab, wie (gut) Entwickler die Kooperation von automatischem System und Operateur unterstützt haben (Kaber et al., 2001).

Dass automatisch unterstützte besser als manuell arbeitende Operateure in der Lage sind, die Anstrengung der Navigatoren zu reduzieren bzw. den Effekt von Anstrengung auf die Leistung zu verändern, kann zwei Gründe haben: Operateure

könnten durch die Aufgabendelegation an die Automation entlastet werden, was in einem erhöhten bzw. dem Erhalt des Situationsbewusstseins resultieren kann. Dadurch könnten die Operateure in der teilautomatisierten Bedingung das Anstrengungsniveau der Navigatoren besser registriert und ggf. kompensierende Eingriffe zur Anstrengungsreduktion getätigt haben. Andererseits kann die Anstrengung der Navigatoren in der teilautomatisierten Bedingung auch durch die insgesamt geringere Anzahl an Hinweisen reduziert worden sein: Der Mittelwert aller (manuell oder automatisch aktivierten) Hinweisklicks von $M = 49,55$ ($SD = 25,30$) war geringer als der Mittelwert aller (manuellen) Klicks in der vorherigen Bedingung ($M = 68,90$, $SD = 43,35$; $T = 2.465$, $p < .05$). Weitere Befunde zeigen, dass Operateure in der automatisch unterstützten Bedingung, wie zu erwarten ist, sehr viel weniger manuelle Hinweise geben, da die Automation diese Aufgabe übernimmt. Das Wirken der Automation erfolgt im Vergleich zu manuellen Hinweisen hoch regelhaft und exakt, indem bspw. Richtungshinweise stets am exakt selben Streckenpunkt und nur vor einer Gabelung gegeben werden. Die rein manuelle Unterstützung wird dagegen sehr viel inkonsistenter erfolgen. Das könnte zu einer höheren Anstrengung der Navigatoren geführt haben. Weitere hier aufschlussgebende Daten werden aktuell ausgewertet.

Offen ist auch, wie sich die Interaktionen des Operateurs mit komplexeren Formen automatischer Unterstützung, bspw. die Zusammenarbeit mit adaptiver Automation auf die Leistung auswirkt. Angestrebt wird, die Simulationsumgebung in Anlehnung an die Level of Automation (Sheridan & Verplank, 1978), iterativ um höhere Automationsstufen zu erweitern und deren jeweiligen Einfluss auf die Leistung zu überprüfen. Dieses Vorgehen stellt einen Vorteil gegenüber Studien (z.B. Sauer et al., 2013) dar, in denen in einer Versuchssituation mehrere Automationsstufen mit mehreren Formen der Kontrolle (z.B. statisch vs. adaptiv) kombiniert werden, die jedoch Gefahr laufen, dass die vielschichtigen Effekte auf die Leistung im Nachhinein nur wenig eindeutig und damit schwer zu erklären sind. Mit dem hier gewählten schrittweisen Vorgehen sind valide Rückschlüsse auf Kriterien für dynamische Funktionsallokation (z.B. Notwendigkeit der Entlastung des Operateurs als Initiator für das adaptive Einschalten von Automaten) möglich.

Literatur

- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Cummings, M. L., & Thornburg, K. (2011). Paying attention to the man behind the curtain. *IEEE Pervasive Computing*, 10(1), 58-62.
- Dekker, S. W. A., & Woods, D. D. (2002). MABA-MABA or abracadabra? Progress on human- automation coordination. *Cognition, Technology, and Work*, 4, 240-244.

- Fitts, P. M. (1951). Human engineering for an effective air navigation and traffic control system. Ohio State University Research Foundation Report. Columbus, OH.
- Flemisch, F., Kelsch, J., Löper, C., Schieben, A., Schindler, J., & Heesen, M. (2008). Cooperative Control and Active Interfaces for Vehicle Assistance and Automation. München: FISITA World automotive Congress.
- Hancock, P.A. (2007). The effects of automation invocation procedure and dynamic display relocation on performance in a multi-task environment. *Transaction of the IEEE on Systems, Man, and Cybernetics, Part A: Humans and Systems*, 37(1), 47-57.
- Hildebrandt, M., & Harrison, M.D. (2002). The temporal dimension of Dynamic Function Allocation. *Proceedings of 11th European Conference on Cognitive Ergonomics*, 283-292.
- Kaber, D., Riley, J., Tan, K., & Endsley, M. (2001). On the design of adaptive automation for complex systems. *Journal of Cognitive Ergonomics*, 5(1), 37-57.
- Kienle, M., Damböck, D., Kelsch, J., Flemisch, F., & Bengler, K. (2009): Towards an H-Mode for highly automated vehicles: Driving with side sticks. In: *Automotive User Interfaces and Interactive Vehicular Applications*. Essen.
- Lemoine, M.P., Debernard, S., & Millot, P. (1995). Men-machines cooperation: toward an experimentation of a multi-level cooperative organization in air-traffic control. In: *IEEE International Conference on System, Man and Cybernetics*, 1503-1508. Vancouver, British Columbia, Canada.
- Norman, D.A. (2007). *The Design of Future Things*. New York: Basic Books.
- Parasuraman, R., & Wickens, C.D. (2008). Humans: Still vital after all these years of automation. *Human Factors*, 50, 511-520.
- Sauer J., Kao C.S., & Wastell D. (2012). A comparison of adaptive and adaptable automation under different levels of environmental stress. *Ergonomics*, 55, 840-853.
- Sauer J, Nickel P, & Wastell D. (2013). Designing automation for complex work environments under different levels of stress. *Applied Ergonomics* 44(1), 119-27.
- Sheridan, T. B., & Verplank, W. (1978). *Human and Computer Control of Undersea Teleoperators*. Cambridge, MA: Man-Machine Systems Laboratory, Department of Mechanical Engineering, MIT.
- Sheridan, T.B., & Parasuraman, R. (2006). Human-automation interaction. *Reviews of Human Factors and Ergonomics*, 1, 89-129.

- Stade, M.J.C., Meyer, C., & Nachtwei, J. (2011). Ergonomische Gestaltung trotz oder aufgrund interdisziplinärer Teamarbeit? 53. *Fachausschusssitzung Anthropotechnik* 27.-28. Oktober 2011, Neu-Isenburg, Deutschland.
- Stade, M., Meyer, C., Niestroj, N., & Nachtwei, J. (2011). (Not) Everybody's Darling: Value and Prospects of Multiple Linear Regression Analysis and Assumption Checking. In: B. Krause, R. Beyer und G. Kaul (Hrsg.), *Empirische Evaluationsmethoden Band 15* (S. 17-34). Berlin: ZeE.
- Wiener, N. (1949). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications* Cambridge MA: MIT Press.
- Zijlstra, F.R.H., & van Doorn, L. (1985). *The construction of a scale to measure perceived effort*. Delft, The Netherlands: Department of Philosophy and Social Sciences, Delft University of Technology.

Partizipative Interaktions- und Automationsgestaltung teil- bis hochautomatisierter Fahrzeuge: Ausgewählte Ergebnisse explorativer Nutzerstudien zu H-Mode 1.1

Sonja Meier¹, Eugen Altendorf¹¹⁹, Marcel Baltzer¹¹⁹ und Frank Flemisch¹¹⁹²

Schlüsselwörter: Mensch-Maschine-Systeme, Mensch-Maschine-Interaktion, partizipative Technikentwicklung, Entwicklungsprozess, Assistenz und Automation

Zusammenfassung

Der folgende Beitrag skizziert die aktuelle Instanz des H-Modes, bei der es sich um eine haptisch-multimodale Bedienweise zum Führen teil- und hochautomatisierter Fahrzeuge auf der Basis der H-Metapher handelt. H-Mode 1.1 besteht im Vergleich zur Version 1.0 (Flemisch et al., 2012) aus den Assistenz- und Automationsgraden assistiert (Tight Rein), teil-/hochautomatisiert (Loose Rein) und temporär vollautomatisiert (Secured Rein). Der Beitrag legt den Schwerpunkt dabei auf den iterativ partizipativen Charakter des Design- und Entwicklungsprozesses der Interaktions- und Automationsgestaltung. Hierzu wurde eine Reihe von projektbegleitenden Nutzerstudien im Fahrsimulator durchgeführt. Die gewonnenen Erkenntnisse fließen direkt in den Design- und Entwicklungsprozess des soziotechnischen Mensch-Maschine-Systems – bestehend aus Fahrer und Automation – ein und führen zu einer hohen Produktakzeptanz in weiteren Entwicklungsphasen. Nach einer kurzen Einführung in die Thematik der kooperativen Fahrzeugführung sowie der Vorstellung der möglichen Ansätze – insbesondere des H-Mode-Ansatzes – werden die Ergebnisse von zwei Versuchsreihen vorgestellt, hinsichtlich des Interaktions- und Automationsdesigns gegenübergestellt sowie daraus Gestaltungshinweise abgeleitet und diskutiert. Zudem offenbaren die Ergebnisse interessante Tendenzen bezogen auf soziodemographische Faktoren und Fahrgewohnheiten für die Fahrzeugführung.

Einführung

Die Ansprüche an Mobilität und räumliche Flexibilität wachsen kontinuierlich. Dadurch werden auch die zu überwindenden Distanzen, z.B. beim Pendeln zum Arbeitsplatz, bei Dienstreisen oder privaten Verpflichtungen, stetig größer. Zugleich steigen im Individualverkehr die Sicherheitsbedürfnisse sowie die Ansprüche an Komfort und Energieeffizienz.

Bereits heute gibt es zahlreiche Assistenzsysteme, beispielsweise für das Abstandhalten (Adaptive Cruise Control, ACC), das Spurhalten (Lane Keeping Assistant System, LKAS) und das Einparken. Der Nutzer sieht sich zunehmend mit ei-

¹ RWTH Aachen, Lehrstuhl und Institut für Arbeitswissenschaft

² Fraunhofer FKIE, Abteilung Ergonomie und Mensch-Maschine-Systeme

ner Reihe von Unterstützungsfunktionen konfrontiert, die nur bedingt einen intuitiven Zugang ermöglichen. Im ungünstigsten Fall können sie sogar zu einem Sicherheitsrisiko führen, da die Aufmerksamkeit des Fahrers von der Hauptfahraufgabe auf ihre Bedienung gelenkt wird oder bei zu hohen Automationsgraden ein "operator-out-of-the-loop"-Problem auftreten kann (z.B. Kelsch et al., 2006).

Kooperative Fahrzeugführung mit dem H-Mode

Der H-Mode als haptisch-multimodale Bedienweise assistierter bis hochautomatisierter Fahrzeuge ist ein ganzheitlicher Ansatz der kooperativen Fahrzeugführung (Flemisch et al., 2006), der durch die H(orse)-Metapher (Flemisch et al., 2003) inspiriert ist. Er orientiert sich an dem natürlichen Vorbild von Pferd und Reiter. Durch das synergetische Zusammenwirken von Mensch und Maschine entsprechend Abbildung 1 entstehen Mehrwerte, die positive Effekte u.a. für die Sicherheit mit sich bringen (Flemisch et al., 2012). Eine weitere Ausprägungsform der kooperativen Fahrzeugführung ist der mehr mit diskreter Beauftragung arbeitende Ansatz Conduct-by-Wire (Winner et al., 2006).

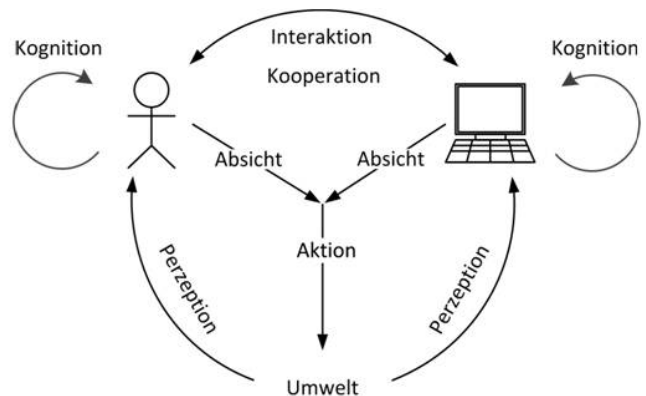


Abb.1: links: Theatertechnik im IAW-SmpLab; rechts: Kooperation von Mensch und Maschine (Flemisch et al., 2012a i.A.a. Flemisch et al. 2006, Kelsch et al. 2006).

Der H-Mode verfolgt eine dynamische Kontrollverteilung zwischen Mensch und Maschine, d.h. die Automation ist in der Lage, weite Teile der Fahrzeugführung zu übernehmen, der Mensch bleibt jedoch stets kooperativ in die Aufgabe der Fahrzeugführung eingebunden (z.B. Schindler et al., 2007). Die Kontrollverteilung kann über einen berührungssensitiven Bildschirm, aber auch fließend zwischen Mensch und Maschine arbitriert bzw. ausgehandelt werden (z.B. Flemisch et al., 2008, vgl. auch Bruder et al., 2011). Die Definitionen von Assistenz- und Automationsgraden befinden sich aktuell in der rechtlichen Diskussion (Gasser et al., 2012). Beim H-Mode sind diese Modi klar definiert und auf einer Assistenz- und Automationsskala abgebildet (Abb. 2). Die Assistenz- und Automationsgrade sind dabei als diskret zu verstehen, die Übergänge (Transitionen) können auch fluide u.a. über die Berücksichtigung der Griffkraft des Fahrers erfolgen (Flemisch et al., 2012).

Ursprünglich waren in Anlehnung an die H-Metapher zwei Modi (Tight Rein, Loose Rein) bei der NASA- und DLR-Instanziierung (Flemisch et al., 2012) auf der Skala angeordnet. Beim Tight Rein (assistiert/niederautomatisiert) behält der Mensch die ausgeprägte Kontrolle über die gemeinsame Bewegung und die Maschine assistiert. Beim Loose Rein (teil-/hochautomatisiert) bestimmt der Mensch die übergeordnete Richtung und bleibt in Kontrolle, die Automation übernimmt jedoch bereits größere Teile der Umsetzung. Wichtig ist hier die Weiterführung kontinuierlicher Kommunikation zwischen Mensch und Automation, um dem "operator-out-of-the-loop"-Risiko vorzubeugen und Rücktransitionen in weniger automatisierte Modi z.B. bei Systemfehlern problemlos durchführen zu können.

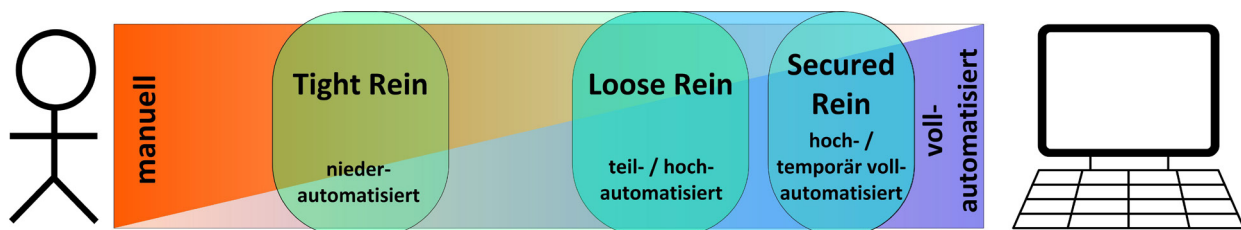


Abb.2: Assistenz- und Automationsskala (i.A.a. Flemisch et al., 2008a).

Bei der Weiterentwicklung des H-Mode 1.1 am IAW der RWTH Aachen wurde ein weiterer Modus integriert, der Secured Rein. Dieser ist stark beeinflusst durch den – in den Projekten HAVEit (Hoeger et al., 2011) und CityMobil (CityMobil, 2012) erarbeiteten – Anwendungsfall des temporär vollautomatisierten Fahrens. Secured Rein steht für hoch- bzw. temporär vollautomatisiertes Fahren, d.h. der Mensch kann zeitlich begrenzt vollständig die Kontrolle an die Automation abgeben. Entscheidend ist hier ein genügend ausgeprägtes Situationsbewusstsein der Automation bezüglich ihrer eigenen Leistungsfähigkeit, um den Fahrer bei Bedarf rechtzeitig wieder in die Fahraufgabe einbinden zu können.

Explorative Nutzerstudien zur Weiterentwicklung des H-Mode 1.1

Parallel zur Entwicklung des H-Modes wurde mit einer Reihe von Partnern eine iterativ partizipative Vorgehensweise entwickelt (z.B. Schindler et al., 2007, Flemisch et al., 2008b, Schieben et al., 2009, Flemisch, 2003), die auch am IAW der RWTH Aachen aufgegriffen und weiterentwickelt wird. Diese Entwicklung hin zum partizipativen Gestalten ist auch in anderen Domänen zu beobachten, wie z.B. zur partizipativen Nutzereinbindung im Bereich der Servicerobotikanwendungen im Kontext stationärer Pflege (Derpmann & Compagna, 2009).

In den hier beschriebenen Simulatorstudien partizipierten 20 Probanden aufgeteilt auf zwei Versuchsreihen (V1, V2). Die erste fand im April 2013 statt und es nahmen zehn Probanden (5 m, 5 w) mit einem Durchschnittsalter von 24.3 Jahren (SD = 3.7 Jahre) teil. Die zweite Reihe im Juli 2013 hatte eine ähnliche demographische Verteilung mit zehn Probanden (5 m, 5 w) und einem Durchschnittsalter von 24.6 Jahren (SD = 4.1 Jahre). Ein aktiver Sidestick wurde als ein mögliches Stellteil neben aktivem Gaspedal und Lenkrad verwendet (Flemisch et al., 2012).

Die Szene bildete ein Streckenabschnitt einer dreispurigen Autobahn mit Fremdfahrzeugen. In der zweiten Versuchsreihe wurde dieser um einen Rundparcours mit einer Landstraße in einem zusätzlichen Teil ergänzt.

Die in diesem Rahmen durchgeführten projektbegleitenden Nutzerstudien umfassten eine Vorbefragung bezüglich soziodemographischer Daten, Fahrgewohnheiten und Persönlichkeitseigenschaften, einen explorativen Teil zur direkten Einbindung der Nutzer in den Designprozess und einen zweistufigen Evaluationsteil, eingeteilt in einen – bezogen auf die haptisch-multimodale Bedienweise des H-Modes – untrainierten und einen trainierten Zustand des Fahrers.

Bei den Probandenstudien wurden ausgewählte Methoden für erwartungsbasierte, partizipative Gestaltung des Interaktions- und Automationsdesigns im Anwendungskontext verwendet. Dazu gehörten z.B. das strukturierte Interview zu erwartetem Design, die (Concurrent-) Think-Aloud-Methode (Häder, 2006) während der Fahraufgaben sowie die Theatertechnik (Schieben et al., 2009). Diese Technik entwickelt den Wizard-of-Oz-Ansatz dahingehend weiter, dass der Nutzer und ein Confederate des Entwicklungsteams, der das technische Untersystem nachbildet, zusammen ein Interaktionsdesign entwickeln bzw. erspielen. In der vorliegenden Studie wurde diese Theatertechnik auf die kooperative Fahrzeugführung dahingehend angewandt, dass der Proband haptisch mit dem Confederate gekoppelt ist. Dieser kann so ein mentales Modell der Vorstellungen und Erwartungen des Probanden aufbauen. Für die Evaluation der aktuellen Instanz des H-Modes wurden Fragebögen z.B. zu Mode Awareness sowie der Beurteilung von Assistenz- und Automationsgraden verwendet. Diese Teile ergänzen sich gewinnbringend für Rückschlüsse auf die Gestaltung kooperativer Mensch-Maschine-Systeme.

Ergebnisse und Diskussion

Insgesamt ließ sich anhand der Ergebnisse erkennen, dass bereits eine kurze Einführung im zehnminütigen Training ausreichte, um eine deutliche bessere Verständlichkeit der H-Mode-Bedienweise zu erzielen. Im Folgenden werden die Ergebnisse des Hotruns – Bewertung im trainierten Zustand des Fahrers – verwendet.

Die Frage nach der Bewertung des H-Mode 1.1 insgesamt beantworteten die Teilnehmer mit „ziemlich gut“ (V1: $M = 5.8$, $SD = 0.6$; V2: $M = 5.8$, $SD = 0.9$; 1 = „sehr schlecht“ bis 7 = „sehr gut“). Die weiteren Aspekte Nützlichkeit, empfundene Kooperation zwischen Fahrer und Automation, Erleichterung des Fahrens, gefühlte Sicherheit sowie angenehmes Fahren sind in Abbildung 3 dargestellt. Die Ergebnisse der beiden Versuchsreihen unterscheiden sich dabei nur unwesentlich und sämtliche Items wurden positiv bewertet.

Interessant ist weiterhin die Betrachtung der Ergebnisse zu den Assistenz- und Automationsgraden. Gemittelt wurde der Loose Rein (teil-/hochautomatisiert) am positivsten bewertet. In Abbildung 4 lässt sich insbesondere beim Loose und beim

Secured Rein jedoch eine jeweils entgegengesetzte Tendenz bei der Bewertung feststellen, deren Signifikanz noch zu untersuchen sein wird.

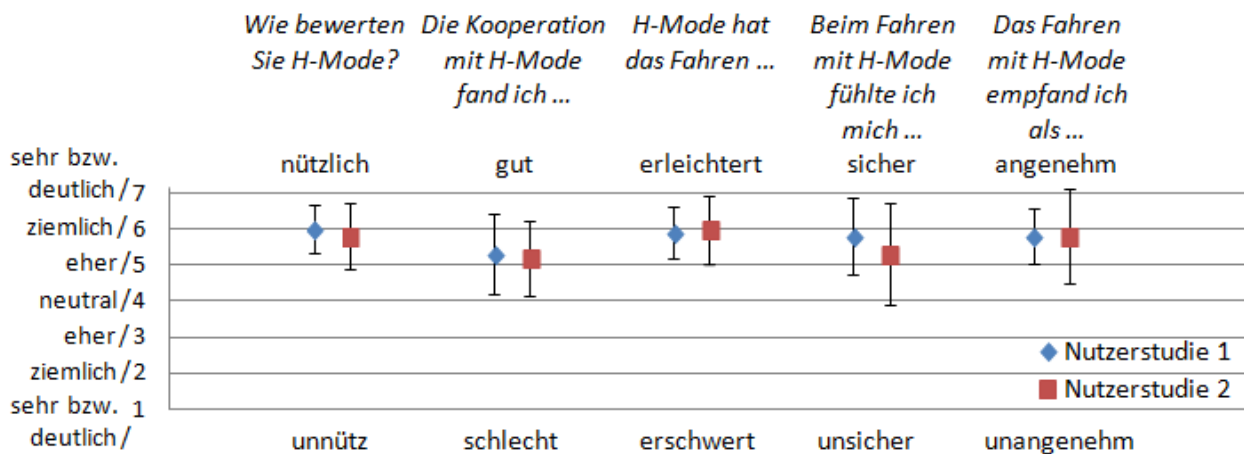


Abb.3: Ausgewählte Aspekte zur Bewertung des H-Mode 1.1 (M±SD).

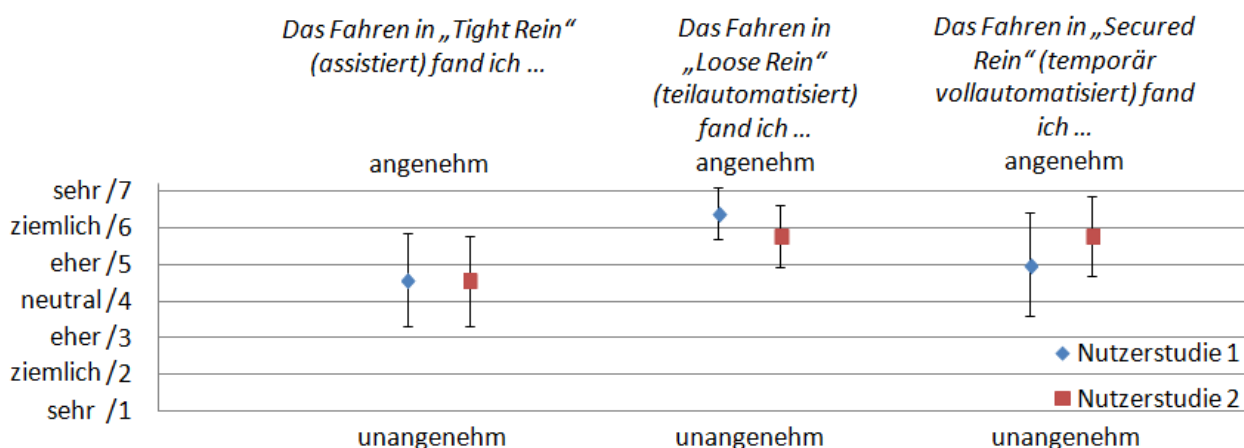


Abb.4: Bewertung des Fahrens in den Assistenz- und Automationsgraden (M±SD).

Der Loose Rein wurde in beiden Versuchsreihen als "ziemlich angenehm" eingeschätzt allerdings mit leicht fallender Tendenz (V1: M = 6.4, SD = 0.66; V2: M = 5.8, SD = 0.8; 1 = "sehr unangenehm" bis 7 = "sehr angenehm"), wohingegen der Secured Rein tendenziell positiver bewertet wurde (V1: M = 5.0, SD = 1.34; V2: M = 5.8, SD = 1.1; 1 = "sehr unangenehm" bis 7 = "sehr angenehm"). Der Grund hierfür ist in einer Designentscheidung zu vermuten. Bei der ersten Versuchsreihe überholte die Automation in keinem Modus selbstständig, was von Probanden kritisiert wurde: "Schlecht fand ich, dass es [das Auto] im Secure[d] Mode nicht selbstständig überholen konnte." Dieses Automationsdesign wurde für die zweite Versuchsreihe daher so angepasst, dass die Automation im Secured Rein selbstständig überholen konnte. Im Loose Rein hingegen war der Überholvorgang immer noch vom Menschen zu initiieren, wenn die Umsetzung auch durch die Maschine unterstützt wurde. Der direkte Vergleich von Loose Rein und Secured Rein führte daher anscheinend dazu, dass der Secured Rein in der zweiten Versuchsreihe ebenso angenehm eingeschätzt wurde wie der Loose Rein, bei dem nun verstärkt der Wunsch auftauchte, auch in diesem ein automationsinitiiertes Überholen zu ermög-

lichen ("Im LR kein Spurwechsel wenn Fahrzeug vorne fährt."). Diese Detailauslegung des Spurwechsels wird in zukünftigen Studien zu untersuchen sein.

Für den fahrerinitiierten Wechsel zwischen den Automationsgraden gab es wie bereits angedeutet zwei Optionen: eine "konventionelle" Auswahl über Buttons auf einem Touchscreen, der vor dem Fahrer angebracht war, sowie im fluiden Modus über Griffkraft direkt am Sidestick. Das subjektive Gefallen bei den Probanden wies zwischen den beiden Optionen keine großen Unterschiede auf. Abbildung 5 zeigt jedoch eine positive Tendenz für die alternativen Auswahlmöglichkeiten. Dies ist höchstwahrscheinlich auf die Position des Touchscreens zurückzuführen. Während sich dieser bei der ersten Versuchsreihe vergleichsweise niedrig ca. auf Kniehöhe befand, wurde er für die zweite Studie stärker ins Zentrum des Blickfeldes gerückt, sodass Probanden tendenziell positivere Bewertungen für die Bildschirmposition in den Interviews der zweiten Versuchsreihe abgaben und dies mit einer geringeren Ablenkung von der Straße begründeten. Für die Einschätzung des Einflusses der Transitionen auf die Kontrollierbarkeit des Fahrzeugs wird angestrebt, eine gezielte Studie zu den Griffkraftschwellen durchzuführen.

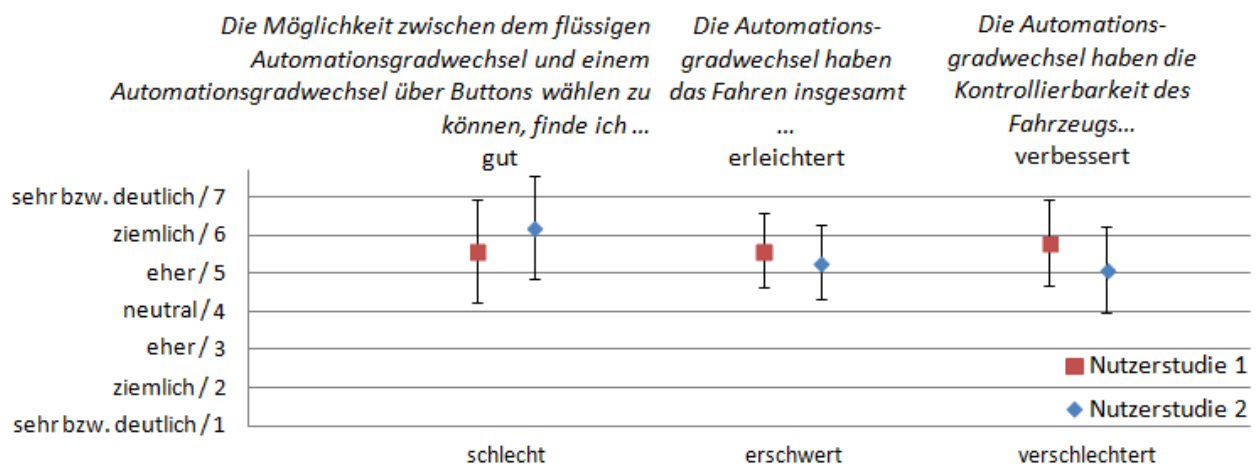


Abb.5: Bewertung der Transitionsoptionen (M±SD).

Die Ergebnisse der beiden Versuchsreihen sind insgesamt bereits vielversprechend. Zugleich unterstreichen sie die Bedeutung der wiederholten Einbindung von Nutzern in den Entwicklungsprozess, um z.B. indirekte Auswirkungen von Designentscheidungen aufzudecken. Durch eine detailliertere Analyse der Ergebnisse und der Detailfragen sind weitere Erkenntnisse zu erwarten. Interessant sind zudem die sich abzeichnenden Tendenzen der soziodemographischen Daten sowie der Fahrgewohnheiten auf die subjektive Wahrnehmung des kooperativen Ansatzes zur Fahrzeugführung. Wenngleich vermutet werden könnte, dass insbesondere "Vielfahrer" und generell technikbegeisterte Personen Gefallen an dem Ansatz finden, zeigen die Ergebnisse eine überraschend positive Wahrnehmung auch und zum Teil insbesondere weniger technikaffinerer Nutzergruppen, was ein großes Potential für derartige Ansätze vermuten lässt. Weitere Studien mit größeren Probandenzahlen werden zur empirischen Absicherung sowie Verfeinerung der Ergebnisse und gewonnenen Erkenntnisse angestrebt.

Literatur

- Bruder, R., Franz, B., Kauer, M., & Schreiber, M. (2011). Fahrerverhalten einer kooperativen Fahrer-Fahrzeug-Interaktion. Zukunft der Fahrzeugführung - kooperativ oder autonom?. 5. Darmstädter Kolloquium Mensch & Fahrzeug, Darmstadt.
- CityMobil (2012). Onlinequelle. URL: <http://www.citymobil-project.eu/index.php>.
- Derpmann, S., & Compagna, D. (2009).: Erste Befunde der Bedarfsanalyse für eine partizipative Technikentwicklung im Bereich stationärer Pflegeeinrichtungen. In: D.iego Compagna, K.aren Shire (Hrsg.),: *Working Papers kultur- und techniksoziologische Studien* (WPktS 05/2009).
- Flemisch, F., Kelsch, J., Schieben, A., & Schindler, J. (2006). Stücke des Puzzles hochautomatisiertes Fahren: H-Metapher und H-Mode, Zwischenbericht. In: Stiller, C., M. Maurer (Hrsg.), 4. *Workshop Fahrerassistenzsysteme*. 60–69. Freundeskreis Mess- und Regelungstechnik Karlsruhe e.V., Karlsruhe.
- Flemisch, F., Kelsch, J., Löper, C., Schieben, A., & J. Schindler, J. (2008). Automation spectrum, inner/outer compatibility and other potentially useful human factors concepts for assistance and automation. In: de Waard, D., F.O. Flemisch, B. Lorenz, H. Oberheid, K.A. Brookhuis (eds.), *Human Factors for assistance and automation* 1-16. Maastricht, Niederlande: Shaker Publishing.
- Flemisch, F., Kelsch, J., Löper, C., Schieben, A., Schindler, J., & Heesen, M. (2008a). Cooperative control and active interfaces for vehicle assistance and automation. *FISITA World automotive Congress*, München.
- Flemisch, F., Meier, S., Baltzer, M., Altendorf, E., Heesen, M., Griesche, S., Weißgerber, T., Kienle, M., & Damböck, D. (2012). Fortschrittliches Anzeige- und Interaktionskonzept für die kooperative Führung hochautomatisierter Fahrzeuge: Ausgewählte Ergebnisse mit H-Mode 2D 1.0. 54. *Fachausschuss-sitzung Anthropotechnik: Fortschrittliche Anzeigesysteme für die Fahrzeug- und Prozessführung*; Koblenz.
- Flemisch, F., Schindler, J., Kelsch, J., Schieben, A., & Damböck, D. (2008b). Some Bridging Methods towards a Balanced Design of Human-Machine Systems, Applied to Highly Automated Vehicles. *Applied Ergonomics International Conference*, Las Vegas, USA.
- Flemisch, F., Meier, S., Neuhöfer, J., Baltzer, M., Altendorf, E., & Özyurt, E. (2012a). Kognitive und kooperative Systeme in der Fahrzeugführung: Selektiver Rückblick über die letzten Dekaden und Spekulation über die Zukunft.; *Workshop Kognitive Systeme*, Duisburg.
- Flemisch, F.O. (2003): Was kommt nach User Centered Design? (Skizze auf der Suche nach Leitmotiven für das Entwerfen und Gestalten). 5. *Berliner Werkstatt Mensch-Maschine Systeme*, Entwerfen und Gestalten, Berlin.
- Flemisch, F.O., Adams, C. A., Conway, S. R., Goodrich, K. H., Palmer, M. T., Schutte, P. C. (2003). *The H-Metaphor as a guideline for vehicle automation*

- and interaction*. NASA/TM—2003-212672, NASA Langley Research Center, Hampton, Va, USA.
- Gasser, T., Arzt, C., Ayoubi, M., Bartels, A., Eier, J., Flemisch, F., Häcker, D., Hesse, T., Huber, W., Lotz, C., Maurer, M., Ruth-Schumacher, S., Schwarz, J., & Vogt, W. (2012). Projektgruppe "Rechtsfolgen zunehmender Fahrzeugautomatisierung", BASt.
- Häder, M. (2006). Empirische Sozialforschung – Eine Einführung, Lehrbuch. Wiesbaden: VS Verlag für Sozialforschung.
- Hoeger, R., Zeng, H., Hoess, A., Kranz, T., Boverie, S., Strauss, M., Jakobsson, E., Beutner, A., Bartels, A., To, T., Stratil, H., Fürstenberg, K., Ahlers, F., Frey, E., Schieben, A., Mosebach, H., Flemisch, F., Dufaux, A., Manetti, D., Amditis, A., Mantzouranis, I., Lepke, H., Szalay, Z., Szabo, B., Luithardt, P., Gutknecht, M., Schoemig, N., Kaussner, A., Nashashibi, F., Resende, P., Resende, Vanholme, B., Glaser, S., Allemann, P., Seglö F., & Nilsson, A. (2011). HAVEit Deliverable D61.1 – Final Report.
- Kelsch, J., Flemisch, F.O., Löper, C. Löper, Schieben, A., & Schindler, J., (2006). Links oder rechts, schneller oder langsamer? Grundlegende Fragestellungen beim Cognitive Systems Engineering von hochautomatisierter Fahrzeugführung. M. Grandt, A. Bauch (Hrsg.), *Cognitive Systems Engineering i.d. Fahrzeug- und Prozessführung*., 48. FAS Anthropotechnik, Karlsruhe.
- Schieben, A., Heesen, M., Schindler, J., Kelsch, J., & Flemisch, F. (2009). The theater-system technique: Agile designing and testing of system behavior and interaction, applied to highly automated vehicles.; *Automotive User Interfaces and Interactive Vehicular Applications* (AutomotiveUI), Essen..
- Schindler, J., & Flemisch, F. (2007). Prospektive Gestaltung von Fahrzeugassistenten, Agile Prototypentwicklung und integriertes Testen am Beispiel des DLR-SmpLab.; 7. *Berliner Werkstatt Mensch-Maschine-Systeme*; Berlin.
- Winner, H., Hakuli, S., Bruder, R., Konigorski, U., & Schiele, B. (2006). Conduct-by-Wire - ein neues Paradigma für die Entwicklung der Fahrerassistenz. 4. *Workshop Fahrerassistenzsysteme*, Löwenstein.

Usability und Sicherheit – Ein Widerspruch oder zwei Seiten einer Medaille?

Katharina Sachse, Manfred Thüring und Josephine Grauert¹

Schlüsselwörter: Usability-Evaluation, Sicherheitsrelevante MMS

Zusammenfassung

Schlechte Usability ist bei sicherheitsrelevanten Mensch-Maschine-Schnittstellen (MMS; z.B. in kerntechnischen Anlagen oder in Verkehrsleitzentralen) nicht nur ein Ärgernis, sondern kann auch zu Bedienfehlern führen, die ein erhebliches Sicherheitsrisiko darstellen. Um diesem Umstand Rechnung zu tragen, wurde ReMUS entwickelt, ein Software-Tool, mit dem die Benutzbarkeit digitaler Bedienschnittstellen unter besonderer Berücksichtigung von Sicherheitsanforderungen bewertet werden kann. Es ermöglicht die Kombination von Nutzerfragebögen mit Gewichtungspituren, so dass Nutzungsprobleme nicht nur identifiziert, sondern auch hinsichtlich ihrer Sicherheitsrelevanz priorisiert werden können. Die Konzeption wurde durch Experteninterviews validiert. Die Umsetzung von ReMUS wurde mittels Nutzertests und Expertenreviews evaluiert. Das Tool erwies sich dabei als geeignetes, nützliches sowie akzeptiertes Instrument zur Bewertung sicherheitsrelevanter digitaler MMS.

Einleitung

Heute wird in vielen Bereichen Software eingesetzt, um sicherheitskritische Prozesse zu steuern oder zu überwachen. Beispiele sind moderne digitale Leitwarten in Industrieanlagen und Kraftwerken sowie Software aus dem Bereich der Medizintechnik. Für Software zum Einsatz in sicherheitskritischen Anwendungsbereichen sind umfangreiche Sicherheitsanforderungen definiert, die dafür sorgen sollen, dass von der Software kein Risiko für Mensch und Umwelt ausgeht. Ein Beispiel sind Normen für rechnerbasierte Systeme zum Einsatz in Kernkraftwerken (DIN IEC 60880, DIN IEC 62138). In ihnen ist geregelt, welche Sicherheitsmaßnahmen bei der Entwicklung, Implementierung und Modifizierung zu erfüllen sind. Die Erfüllung solcher Anforderungen wird durch Tests validiert. Während sogar die Anforderungen an diese Tests in den Normen definiert sind, wird auf die Passung von Bediener und Maschine für rechnerbasierte Systeme in Kernkraftwerken erstaunlicherweise nicht eingegangen. Es gibt weder konkrete Hinweise zur sicheren Gestaltung der MMS noch die Forderung, diese zu prüfen. Die mangelnde Berücksichtigung ergonomischer Erfordernisse kann zu schlecht gestalteten Benutzerschnittstellen führen, die Bedienfehlern provozieren können, welche in sicher-

¹ Technische Universität Berlin, Kognitionspsychologie und Kognitive Ergonomie

heitsrelevanten Anwendungsbereichen schwerwiegende Konsequenzen haben können. (Leveson 1995).

In der Medizintechnik ist die Berücksichtigung ergonomischer Anforderungen hingegen bereits explizit geregelt. Durch die DIN Normen EN 62366 und EN 60601-1-6 ist definiert, wie bei der Entwicklung medizintechnischer Hardware und Software ein Usability Engineering Prozess geplant, durchgeführt und dokumentiert werden muss. Der Benutzer wird also von Anfang an bei der Systementwicklung berücksichtigt. Durch die gebrauchstaugliche Gestaltung der Bedienschnittstellen sollen Bedienfehler vermieden und die Benutzer entlastet werden. Beides kann einen positiven Effekt auf die Sicherheit der Patientenversorgung haben.

Die Beispiele Kerntechnik und Medizintechnik zeigen erhebliche Unterschiede, wie mit dem Thema Gebrauchstauglichkeit von sicherheitsrelevanten Technologien umgegangen wird. In der Medizintechnik gilt gute Usability als Voraussetzung für die sichere Bedienung von Systemen. In der Kerntechnik (und auch in anderen Bereichen, wie z.B. dem Schienenverkehr) gilt hingegen die Annahme, dass Sicherheit durch Automatisierung gewährleistet wird. Die Gestaltung der MMS folgt den Erfordernissen der Technik. Sichere Bedienung soll durch Ausbildung und Training gewährleistet werden. Usability-Überlegungen spielen bei der Systemgestaltung keine Rolle und in einigen Fällen stehen sogar ergonomische Anforderungen im Widerspruch zu Sicherheitsstandards (Sachse & Thüring 2011).

Das Ziel der hier vorgestellten Arbeiten war es, Möglichkeiten der Usability-Bewertung zu entwickeln, mit denen die Gebrauchstauglichkeit von MMS auch in den Bereichen geprüft werden kann, in denen bisher keine Usability-Tests stattfinden. Dazu muss deren speziellen Voraussetzungen Rechnung getragen werden. Zum einen gilt es, mögliche Konflikte zwischen Sicherheitsmaßnahmen und Usability-Aspekten zu berücksichtigen. Zum anderen sollten die Tests mit möglichst geringem Aufwand durchführbar sein, denn schließlich sind Hersteller und Betreiber (bisher) nicht verpflichtet, die Gebrauchstauglichkeit nachzuweisen.

In einem ersten Schritt wurden Interviews mit Experten nukleartechnischer Systeme geführt, um einen Einblick in die Gestaltung und Nutzung von MMS in diesem Bereich zu gewinnen (Abschnitt 2). Die Ergebnisse flossen in die Entwicklung eines rechnerbasierten Tools (ReMUS) ein, mit dem nutzerzentrierte Usability-Evaluationen durchgeführt werden können (Abschnitt 3). Die Validierung des Tools erfolgte durch Befragungen von Operateuren in kerntechnischen Anlagen (Abschnitt 4). Diese Ergebnisse sowie die Resultate einer Expertenevaluation der Usability wurden zur Verbesserung und Erweiterung von ReMUS genutzt.

Experteninterviews

Um ein Usability-Evaluationsverfahren zu entwickeln, das von den Entscheidern in der Zieldomäne (hier kerntechnische Anlagen) akzeptiert und genutzt wird, ist es nötig, die besonderen Bedingungen dieser Domäne zu identifizieren und zu berücksichtigen. Um Informationen hierüber zu gewinnen, wurden Interviews mit

Experten für digitale Systeme für nukleartechnische Anwendungen geführt. Ziel dieser Interviews war es, spezielle Sicherheitsmaßnahmen von digitalen MMS in kerntechnischen Anwendungen zu identifizieren und zu ermitteln, ob diese einen Einfluss auf die Gebrauchstauglichkeit der Schnittstelle haben. Dazu wurden im Rahmen der Interviews die zentralen Anforderungen an die Gebrauchstauglichkeit interaktiver Software vorgestellt und mit den Experten diskutiert.

Um möglichst unterschiedliche Perspektiven zu erfassen, wurden sowohl zwei Experten eines Herstellers von Software für reaktorphysikalische Anwendungen als auch zwei Experten von Reaktoranlagenbetreibern befragt. Alle besprochenen Systeme (Informations- und Steuerungssysteme) verfügen über grafische Benutzeroberflächen. Die Systeme wurden im Rahmen der Interviews durch die Experten vorgeführt. Die Interviews erfolgten halb-standardisiert an den Arbeitsstätten der Experten. Die Gesprächszeit betrug ca. drei Stunden je Interview.

Für die Analyse der Antworten wurden die von den Experten genannten Sicherheitsmaßnahmen expliziert und zusammengefasst. Durch eine Zuordnung der so identifizierten Maßnahmen zu den Anforderungen an die Gebrauchstauglichkeit wurden mögliche Konflikte zwischen Sicherheit und Usability identifiziert.

Die Ergebnisse zeigen, dass die meisten Sicherheitsmaßnahmen über Gestaltungsmaßnahmen realisiert werden, die gleichzeitig eine hohe Gebrauchstauglichkeit gewährleisten. So ist der Funktionsumfang auf die Aufgaben der jeweiligen Nutzergruppe (Operator, Anlagenphysiker oder Administrator) zugeschnitten. Dies dient zum einen der Sicherheit, da Benutzer keine Handlungen vornehmen können, für die sie keine Berechtigung haben. Zum anderen verbessert es die Effizienz der Bedienung, da verhindert wird, dass nicht benötigte Funktionen versehentlich gewählt werden, so dass es zu Fehlern oder Verzögerungen kommt.

Manche Sicherheitsmaßnahmen konfliktieren jedoch mit Usability-Prinzipien. Ein Beispiel hierfür ist, dass viele Operationen aus Sicherheitsgründen doppelt bestätigt werden müssen. Dies reduziert die Bearbeitungsgeschwindigkeit und unter Umständen die Zufriedenheit der Nutzer. Auch die Möglichkeit, die Benutzeroberfläche an individuelle Vorlieben oder Bedürfnisse anzupassen, ist aus Sicherheitsgründen eingeschränkt. In der Regel arbeiten mehrere Operateure an derselben Schnittstelle. Damit sie sich schnell und zuverlässig orientieren können, sind Änderungen der Darstellung nicht möglich.

Als zentrales Ziel der Oberflächengestaltung wurde von den Experten die Vermeidung von Bedienfehlern genannt. Dies soll meist durch Gestaltungsmaßnahmen erreicht werden, die die Handlungsmöglichkeiten der Bediener stark vorgeben oder einschränken. So werden beispielsweise Plausibilitätschecks für eingegebene Werte durchgeführt und diese nur ins System übernommen, wenn sie in einem zulässigen Bereich liegen. Aus Usability-Sicht ist es hingegen vorteilhaft, wenn Systeme trotz fehlerhafter Bedienung eine erfolgreiche Aufgabenbearbeitung zulassen (Fehlertoleranz). Dies scheint auf den ersten Blick ein Widerspruch zu sein. Wesentlich für eine sichere Bedienung ist jedoch, dass im Ergebnis der Operation keine Fehler

auftreten. Dazu dienen sowohl Maßnahmen, die Bedienfehler von vornherein vermeiden, als auch Möglichkeiten, begangene Fehler einfach und schnell zu beheben.

Zusammenfassend zeigt sich, dass es bei der Gestaltung von MMS in der Kerntechnik mehr Gemeinsamkeiten als Unterschiede zwischen Sicherheitsmaßnahmen und Usability-Anforderungen gibt. Die Befürchtung, dass eine gebrauchstaugliche Oberflächengestaltung im Widerspruch zu einer sicheren Bedienung steht, erscheint daher ungerechtfertigt. Vielmehr reduziert eine gut gestaltete Benutzeroberfläche die Wahrscheinlichkeit von Bedienfehlern und trägt damit unmittelbar zur Sicherheit bei. Um den Sicherheitsaspekt bei der Bewertung der Gebrauchstauglichkeit zu betonen, sollten Usability-Aspekte hinsichtlich ihrer Sicherheitsrelevanz priorisiert werden. So lassen sich jene Aspekte identifizieren, die für die sichere Bedienung besonders wichtig sind.

Das ReMUS Tool

Die Ergebnisse der Interviews zeigen, dass Maßnahmen, die eine sichere Bedienung von digitalen Systemen gewährleisten sollen, eng mit den Standards zusammenhängen, die für eine gute Gebrauchstauglichkeit stehen. Der Einsatz standardisierter Instrumente (Fragebögen), mit denen diese Usability-Standards erfasst werden, erscheint daher auch im Bereich sicherheitsrelevanter Systeme angebracht. Um den Aspekt der Sicherheit im Zuge der Betrachtung der Gebrauchstauglichkeit angemessen bewerten zu können, wurde ein Tool zum Rechnerbasierten Multiattributen Usability-Scoring (ReMUS) entwickelt. Dieses Programm ermöglicht die Überprüfung der Gebrauchstauglichkeit sicherheitsrelevanter Systeme, indem es standardisierte Fragebögen mit Gewichtungsprozeduren kombiniert.

ReMUS umfasst verschiedene etablierte Fragebögen. Zur Zeit sind die Fragebögen AttrakDiff (Hassenzahl et al. 2003), Computer System Usability Questionnaire (CSUQ) (Lewis 1993), IsoMetrics (Gediga et al. 1999), ISONORM 9241-110/S (Prümper in Vorb.), System Usability Scale (SUS) (Brooke 1996), User Experience Questionnaire (UEQ) (Laugwitz et al. 2008), sowie der VisAWI (Moshagen & Thielsch 2012) im Umfang der Datenbank enthalten. Darüber hinaus bietet ReMUS die Möglichkeit, selbst erstellte Fragebögen einzubinden und zu nutzen.

Die durch die Nutzerfragebögen erfassten Gestaltungsanforderungen können in verschiedenen Nutzungssituationen eine unterschiedliche Relevanz besitzen. Besonders für sicherheitsrelevante Systeme sind manche Kriterien anders zu bewerten als für Systeme im alltäglichen Gebrauch. ReMUS bietet deshalb die Möglichkeit, einzelne Standardfragebögen mit Gewichtungsprozeduren zu kombinieren, um so die Relevanz einzelner Aspekte bestimmen zu können. Nutzer füllen dabei zunächst den Fragebogen aus und bewerten anschließend die einzelnen Aspekte hinsichtlich ihrer Wichtigkeit.

Aktuell sind zwei Gewichtungsverfahren implementiert: Relative Weighting und der Analytisch Hierarchische Prozess (AHP) von Saaty (1990). Beide Verfahren dienen der relativen Gewichtung, d.h. die Wichtigkeit einzelner Aspekte wird in

Bezug zur Wichtigkeit aller anderen Aspekte gesetzt. Innerhalb des Relative Weighting Verfahrens werden die Aspekte zuerst entsprechend ihrer Wichtigkeit in eine Rangordnung gebracht. Anschließend werden 100 Punkte auf die Aspekte verteilt, wobei die Punktunterschiede die Wichtigkeitsunterschiede abbilden. Beim AHP hingegen werden alle Aspekte paarweise miteinander verglichen. Über eine Ratingskala wird angegeben, wie wichtig der eine Aspekt im Vergleich zu dem jeweils anderen ist. Aus diesen Angaben werden Gewichte berechnet. Die Gewichte beider Verfahren sind so normiert, dass ihre Summe 1 ergibt.

Durch die Kombination der Gewichte mit den mittels Fragebogen erhobenen Usability-Bewertungen kann ein Usability-Score bestimmt werden, der ausdrückt, inwiefern ein System die für die spezielle Anwendungssituation besonders relevanten Usability-Anforderungen erfüllt.

Um mittels ReMUS eine Usability-Evaluation durchzuführen, kann der Testleiter die vorhandenen Standardfragebögen auswählen und sie beliebig miteinander oder mit selbst erstellten Fragebögen kombinieren. Für die Standardfragebögen sind auch Standardinstruktionen enthalten. Um ein Usability-Projekt an die Besonderheiten des Systems, der Nutzer sowie des Unternehmens anzupassen, können eigene Texte und Abbildungen in die Befragung integriert werden. Der Testleiter kann zusätzlich wählen, ob einzelne Fragebögen mit einer Gewichtungsprozedur verknüpft werden sollen. Außerdem besteht die Möglichkeit, das Befragungsprojekt für eine Offline-Durchführung verfügbar zu machen. Dies erlaubt die Datenerhebung auch an Orten, an denen keine Internetverbindung möglich oder erwünscht ist (z.B. innerhalb von Kraftwerken). Abbildung 1 zeigt die Testleiter-Ansicht in ReMUS, die für die Projektkonfiguration genutzt wird.

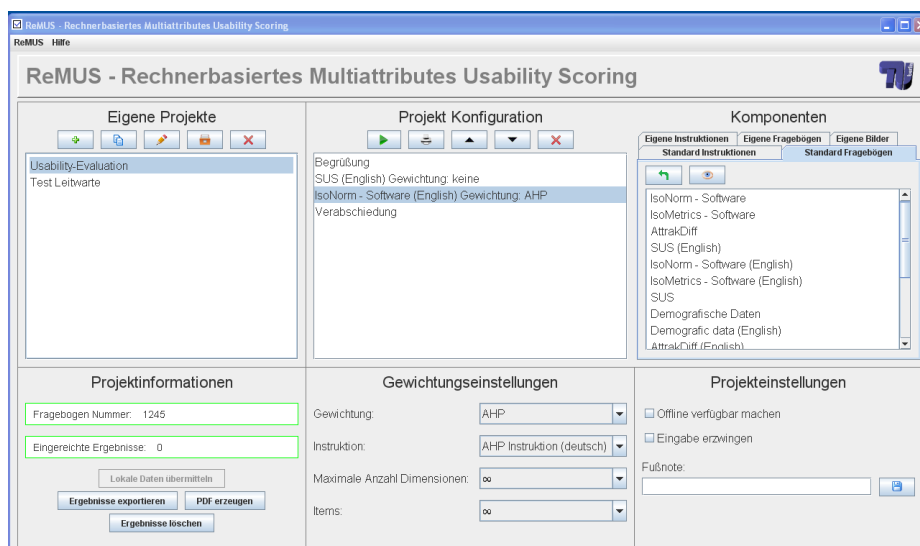


Abbildung 1: ReMUS Screenshot der Testleiteransicht

In der Probandenansicht kann der Nutzer die Fragebögen am Computer bearbeiten. ReMUS speichert die Eingaben und bietet die Möglichkeit, diese nach Ende der Datenerhebung als csv-Datei zu exportieren. Die Exportfunktion bietet zwei

Versionen: einen Basisbericht und einen ausführlichen Bericht. Der Basisbericht enthält die nach den Auswertungsregeln der Standardfragebögen berechneten Ergebniswerte sowie die Gewichte und Usability-Scores für jeden Probanden. Zusätzlich werden Mittelwerte und Standardabweichungen angegeben. Der ausführliche Bericht beinhaltet neben den Daten des Basisberichts zudem sämtliche Rohdaten und ist somit auch für detaillierte Analysen zur Prüfung wissenschaftlicher Fragestellungen geeignet.

Außerdem bietet ReMUS die Möglichkeit, einen Ergebnisbericht im pdf-Format zu erzeugen. Dieser enthält neben einer Kurzbeschreibung der verwendeten Standardverfahren Übersichtstabellen mit den aggregierten Ergebnissen (Mittelwerte und Standardabweichungen) sowie Ergebnisgrafiken. Dieser Bericht ist vor allem für Anwender aus der Praxis geeignet. Abbildung 2 zeigt einen Auszug eines exemplarischen Ergebnisberichtes, der die mittleren Ratings und Gewichte für die verschiedenen Anforderungsdimensionen zeigt.

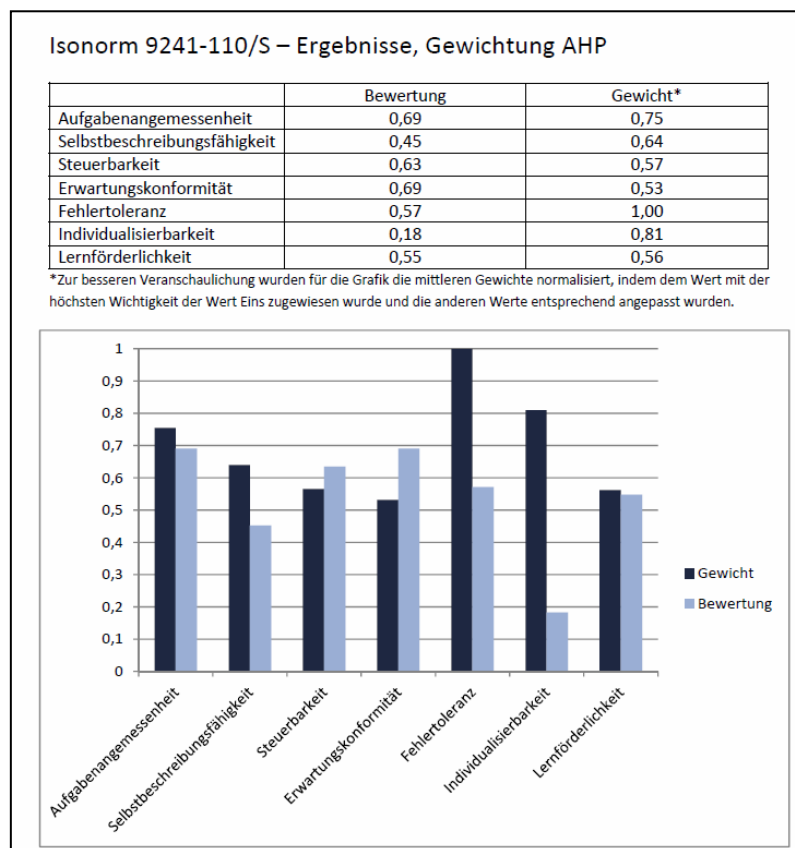


Abbildung 2: Auszug aus dem automatischen Ergebnisbericht

Validierung des Tools in der Praxis

Um die Praxistauglichkeit von ReMUS nachzuweisen, wurden zwei Nutzertests mit Bedienern sicherheitsrelevanter Systeme durchgeführt. Am ersten Test nahmen sechs Operateure eines Forschungsreaktors teil und bewerteten die Gebrauchstauglichkeit der digitalen Leitwarte. In einem zweiten Test bewerteten sieben Anla-

genwärter eines Kernkraftwerkes zwei digitale Schnittstellen zur Bedienung von Hilfssystemen. Die Tests wurden an den Arbeitsstätten der Befragten durchgeführt.

Dabei wurde ReMUS auf einem Notebook präsentiert, an dem die Befragten einzeln nacheinander die Befragung absolvierten. Die Prozedur dauerte je Proband etwa 60 Minuten. Nach einer kurzen mündlichen Einleitung über die Ziele der Untersuchung erfolgte die Bearbeitung der Befragung komplett am Rechner. Eine Versuchsleiterin war jedoch während der Bearbeitung anwesend, um eventuell auftretende Fragen zu beantworten. Sie beobachtete das Verhalten der Befragten und dokumentierte Nutzungsprobleme. Die Befragung startete mit der Aufgabe, typische Arbeitsschritte stichpunktartig zu beschreiben. Dadurch sollte erreicht werden, dass sich die Befragten ihre Interaktion mit dem System vor Augen führen, bevor sie dessen Gestaltung beurteilen. Die Beurteilung der Gebrauchstauglichkeit erfolgte anhand des Fragebogens ISONORM 9241-110/S (Prümper, in Vorb.) Anschließend bewerteten die Befragten die Wichtigkeit der verschiedenen Usability-Aspekte hinsichtlich einer erfolgreichen und sicheren Bedienung. Für die Befragung der Operateure des Forschungsreaktors wurde die Gewichtungsprozedur AHP genutzt. AHP wurde zunächst auch für die Befragung der Anlagenwärter des Kraftwerkes eingesetzt. Es zeigte sich jedoch, dass ein Teil der Befragten Schwierigkeiten mit dieser komplexen Prozedur hat. Aus diesem Grund wurde die Gewichtungsprozedur nach den ersten Durchgängen geändert. Die übrigen Befragten gewichteten die Usability-Aspekte mittels des Relative Weighting Verfahrens.

Die Nutzertests zeigen, dass Operateure bzw. Anlagenwärter von kerntechnischen Anlagen in der Lage sind, die Bedienschnittstellen ihrer Anlage mittels ReMUS zu bewerten, auch wenn einige Befragte Probleme bei einzelnen Schritten der Evaluationsprozedur (AHP-Gewichtung) hatten. Auf diese identifizierten Nutzungsprobleme wurde bei der Weiterentwicklung von ReMUS besonderes Augenmerk gelegt. Es wurde eine vereinfachte Gewichtungsprozedur für AHP implementiert, die die Anzahl der notwendigen Paarvergleiche erheblich reduziert. Auch die Versuchsleiteransicht wurde auf Basis von Ergebnissen einer heuristischen Evaluation optimiert.

ReMUS hat sich in ersten Praxistests bewährt. Die Reaktion auf die Nutzertests war sehr positiv. Die mittels ReMUS erzielten Ergebnisse wurden durch die Verantwortlichen als nützlich und interessant beurteilt, der Aufwand für die Evaluation wurde als eher gering angesehen. ReMUS steht also für den Einsatz in sicherheitsrelevanten Anwendungsbereichen bereit, kann aber auch für die Usability-Evaluation beliebiger anderer MMS genutzt werden.

Literatur

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In Jordan, P. W., Thomas, B., Weerdmeester, B. A. & McClelland, A. L. (Eds.): *Usability Evaluation in Industry*. London: Taylor and Francis. 189-194.

- DIN EN 60601-1-6 (2012). Medizinische elektrische Geräte – Teil 1-6: Allgemeine Festlegungen für die Sicherheit einschließlich der wesentlichen Leistungsmerkmale - Ergänzungsnorm: Gebrauchstauglichkeit. Berlin: Beuth.
- DIN EN 62138 (2010). Kernkraftwerke – Leittechnik für Systeme mit sicherheitstechnischer Bedeutung. Softwareaspekte für rechnerbasierte Systeme zur Realisierung von Funktionen der Kategorie B und C. Berlin: VDE.
- DIN EN 62366 (2008). Medizinprodukte - Anwendung der Gebrauchstauglichkeit auf Medizinprodukte. Berlin: Beuth.
- DIN IEC 60880 (2006). Kernkraftwerke – Leittechnik für Systeme mit sicherheitstechnischer Bedeutung. Softwareaspekte für rechnerbasierte Systeme zur Realisierung von Funktionen der Kategorie A. Berlin: VDE.
- DIN EN ISO 9241-110 (2006). Ergonomics of human-system interaction – Part 110: Dialogue principles. Berlin: Beuth.
- Gediga, G., Hamborg, K.-C. & Düntsch, I. (1999): The IsoMetrics usability inventory: An operationalization of ISO 9241-10 supporting summative and formative evaluation of software systems. *Behaviour & Information Technology*, 151-164.
- Hassenzahl, M., Burmester, M. & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In G. Szwillus & J. Ziegler (Hrsg.), *Mensch & Computer 2003: Interaktion in Bewegung*. Stuttgart: B. G. Teubner. 187-196.
- Laugwitz, B., Held, T. & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In A. Holzner (Hrsg.), *USAB 2008*, LNCS 5298. Springer: Berlin. 125-134.
- Leveson, N.G. (1995). *Safeware, system safety and computers*. Amsterdam: Addison-Wesley.
- Lewis, J.R. (1993). Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. IBM Technical Report 54.786.
- Moshagen, M. & Thielsch, M. T. (2012). A short version of the visual aesthetics of websites inventory. *Behaviour & Information Technology*.
- Prümper, J. (in Vorb.). Der ISONORM 9241/110-S: Kurzfragebogen zur Software-Ergonomie. <http://people.f3.htw-berlin.de/Professoren/Pruemper/instrumente/ISONORM%209241-110-S.pdf>
- Sachse, K. & Thüning, M. (2011). Usability Evaluation für sicherheitsrelevante Software. In: S. Schmid, M. Elepfandt, J. Adenauer & A. Lichtenstein (Hrsg.). *Fortschritt-Berichte VDI, Reflexionen und Visionen der Mensch-Maschine-Interaktion – Aus der Vergangenheit lernen, Zukunft gestalten*. Düsseldorf: VDI Verlag, 365-368.

Förderung

Die Entwicklung von ReMUS wurde gefördert durch das Bundesministerium für Wirtschaft und Technologie, Förderkennzeichen 1501388.

Validierung der Gebrauchstauglichkeit medizintechnischer Produkte mit Hilfe der summativen Methode iFlow

Stefan Pfeffer¹ und Thomas Maier¹

Schlüsselwörter: Usability, Medizinprodukte, Validierung, Quantifizierung, Dokumentation

Zusammenfassung

Die Validierung von Medizinprodukten, die auch in komplexen Mensch-Maschine Systemen eingesetzt werden, stellt eine sehr aufwändige, aber auch notwendige Maßnahme zur Sicherstellung der Gebrauchstauglichkeit dar. Um die Validierung möglichst realitätsnah durchführen zu können, müssen einige Rahmenbedingungen beachtet werden, welche den Einsatz der meisten Usabilitymethoden ausschließen. Die hier vorgestellte iFlow Methode soll eine Möglichkeit bieten, um video- und audiographierte Simulationen zu transkribieren und die darin stattfindenden Mensch-Maschine-Interaktionen hinsichtlich der Usabilitymaße der Effektivität, Effizienz und Zufriedenheit zu bewerten und zu dokumentieren und schlussendlich die Gebrauchstauglichkeit des Medizinproduktes zu validieren. Anhand einer exemplarischen Validierung mit iFlow sollen die Methode vorgestellt und deren Grenzen und Potentiale diskutiert werden.

Einleitung

Für das Inverkehrbringen von Medizinprodukten muss gemäß der Norm DIN EN 62366 (DIN 2011) ein dokumentierter Risikomanagementprozess vorgelegt werden, welcher einen gebrauchstauglichkeitsorientierten Entwicklungsprozess (engl. Usability Engineering Process) beinhalten muss. Hierzu gehören die Aufstellung der Hauptbedienfunktionen (häufig benutzte und sicherheitsrelevante Funktionen), die Spezifikation der Anwendung sowie die daraus resultierende Spezifikation der Gebrauchstauglichkeit (vgl. Abb.1). Auf Basis dieser produktbezogenen Daten wird der Validierungsplan für die Validierung erstellt. Neben den Methoden und Benutzerprofilen für die Validierung werden darin auch die Akzeptanzkriterien, bei denen das Produkt die notwendige Usability erbringt, festgelegt. Verifizierungsverfahren mit formativen Methoden finden iterativ im Sinne einer Zielevaluation während der Produktentwicklung statt. Hierfür stehen gut etablierte Inspektionsmethoden der Usability bereit (Sarodnick&Brau 2011). Diese Methoden eignen sich jedoch nur in geringem Maße für eine umfassende Ergebnisevaluation im Rahmen des Validierungsprozesses. Hier finden summative Methoden Einsatz. Untersuchungen in realen oder quasi-realen (simulierten) Kontexten stellen dabei

¹ Universität Stuttgart, Institut für Konstruktionstechnik und Technisches Design

zusätzlich eine besondere Herausforderungen an die Methodik und die Methoden dar (Morales et al. 2012).



Abb.1: Gebrauchstauglichkeitsorientierter Entwicklungsprozess

Die Durchführung der Validierung erfolgt erst nach erfolgreichem Abschluss der Verifizierung. Wohingegen die Verifizierung z.B. noch im Labor stattfinden kann, sollte die Validierung in einem möglichst realen Kontext der intendierten Anwendung des Produktes stattfinden. Hierzu zählen die Mitbetrachtung von umgebungsbedingten Störvariablen und weiteren Schnittstellen zu Geräten und Personen, sowie der Einfluss von situativen Aspekten wie Zeit- oder Leistungsdruck. Ein System, welches die oben beschriebenen Kriterien beinhaltet, wird auch als komplexes Mensch-Maschine-System bezeichnet.

Die iFlow Methode

Um die Rahmenbedingungen für eine Validierung eines Medizinproduktes, welches in komplexen Mensch-Maschine-Systemen eingesetzt wird, zu erfüllen, ist die Anwendung vieler gängiger Usability-Methoden nicht geeignet. Es darf z.B. keine Ablenkung des Benutzers durch messtechnisches Equipment stattfinden. Ebenso sollten sich keine weiteren Personen (Observatoren, Protokollanten etc.) im Testraum befinden. Die Gebrauchstauglichkeit kann nur dann gewährleistet werden, wenn die Akzeptanzkriterien auch extremen Situationen standhalten. Diese können in Simulationen (quasi-realen Kontexten) hergestellt werden. Für die Datenerhebung eignet sich in diesem Fall nur die Video- und Audiographie. Mithilfe der iFlow Methode („information flow“) können diese Daten quantitativ ausgewertet und visualisiert werden. Die Auswertung findet dabei über zuvor gewählte Interaktionsparameter statt. Die Visualisierung der Mensch-Maschine-Interaktion erfolgt im sog. iFlow-Diagramm. Der Einsatz der iFlow Methode bietet sich vor allem für den zur Validierung in komplexen Mensch-Maschine-Systemen be-

schriebenen Beobachtungskontext an (vgl. Abb.2). Die Bedienpersonen sollten dabei schon mit dem Medizinprodukt vertraut gemacht worden sein und ein explizit ausgewähltes Nutzerprofil aufweisen. Studien zur intuitiven Bedienung, auch mit einer großen Anzahl von allgemeinen Personengruppen, können Teil des Verifizierungsprozesses sein.

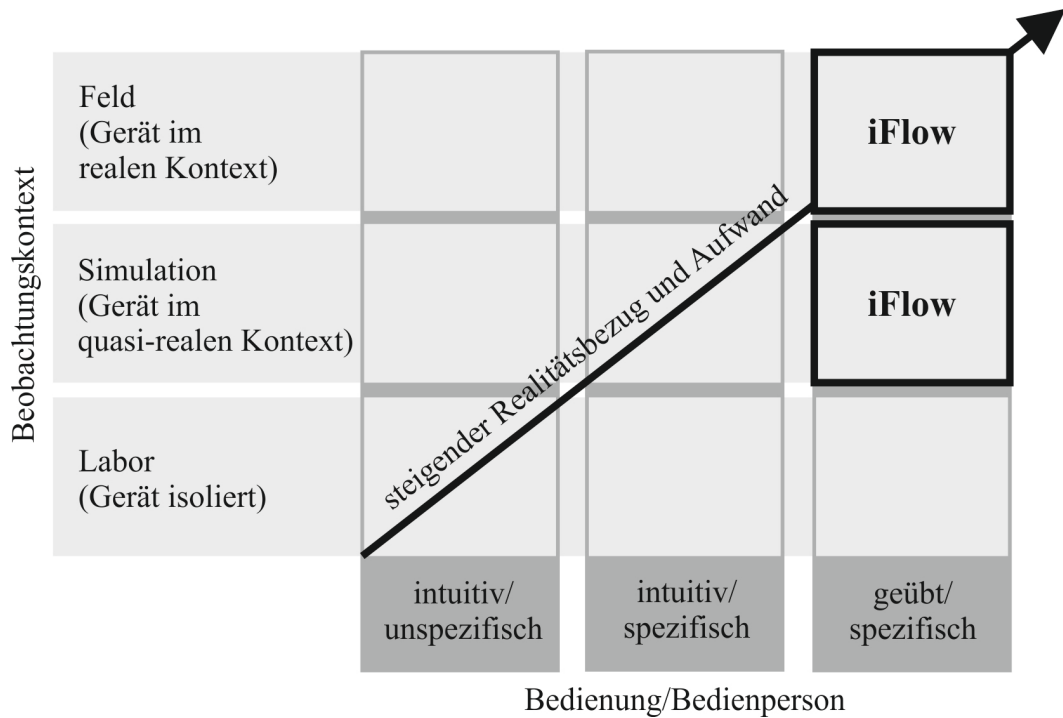


Abb.2: Einsatzbereiche der iFlow Methode

Die Durchführung einer Validierung mit iFlow erfolgt in 5 Schritten (vgl. Abb.3). Im ersten Schritt wird für die Erhebung der Akzeptanzkriterien eine expertenbasierte Bedienanalyse durchgeführt. Hierbei werden alle zuvor definierten Hauptbedienfunktionen im isolierten Kontext unter idealen Bedingungen getestet, die Abläufe video- und audiographiert und in ein separates iFlow-Diagramm (Abschnitt Mensch-Maschine) transkribiert (vgl. Abb.3, Schritt 1). Hieraus werden die Akzeptanzkriterien (Soll-Werte) für die Validierung erhoben, welche sich auf die Usabilitymaße der Effizienz (Anzahl der Bedienschritte bzw. Bedienzeit) und der Effektivität (notwendige Einzelbedienhandlungen und deren Permutation) beziehen. Im nächsten Schritt wird die eigentliche Simulation durchgeführt und diejenigen Sequenzen in das eigentliche iFlow-Diagramm (Abschnitt Mensch-Maschine-Kontext) transkribiert, welche Hauptbedienfunktionen mit dem Medizinprodukt beinhalten (vgl. Abb.3, Schritt 2). Schritt 3 stellt den ersten Auswertungsprozess dar. Hierin werden die Usabilitymaße der Effektivität und Effizienz in Prozent angegeben. In Abbildung 3 wurde Hauptbedienfunktion A (HBF A) aufgrund einer ablenkungsbedingten (Kontext-Schnittstellen) Bedienzeitverlängerung fiktiv mit 40% gewertet, bei 100% Zielerreichung, d.h. bei keinen ausgelassenen bzw. fehlerhaften/unnötigen Bedienhandlungen (Effektivität). Hauptbedienfunktion B (HBF B) entspräche der aufgenommen Soll-Bedienzeit (100% Effizienz), die je-

doch aufgrund ausgelassener oder vergessener Bedienschritte (60% Effektivität) zustande gekommen ist. Das Beispiel zeigt die Relevanz der Betrachtung beider Usabilitymaße, da eine Einzelbetrachtung zu Fehlschlüssen führen kann.

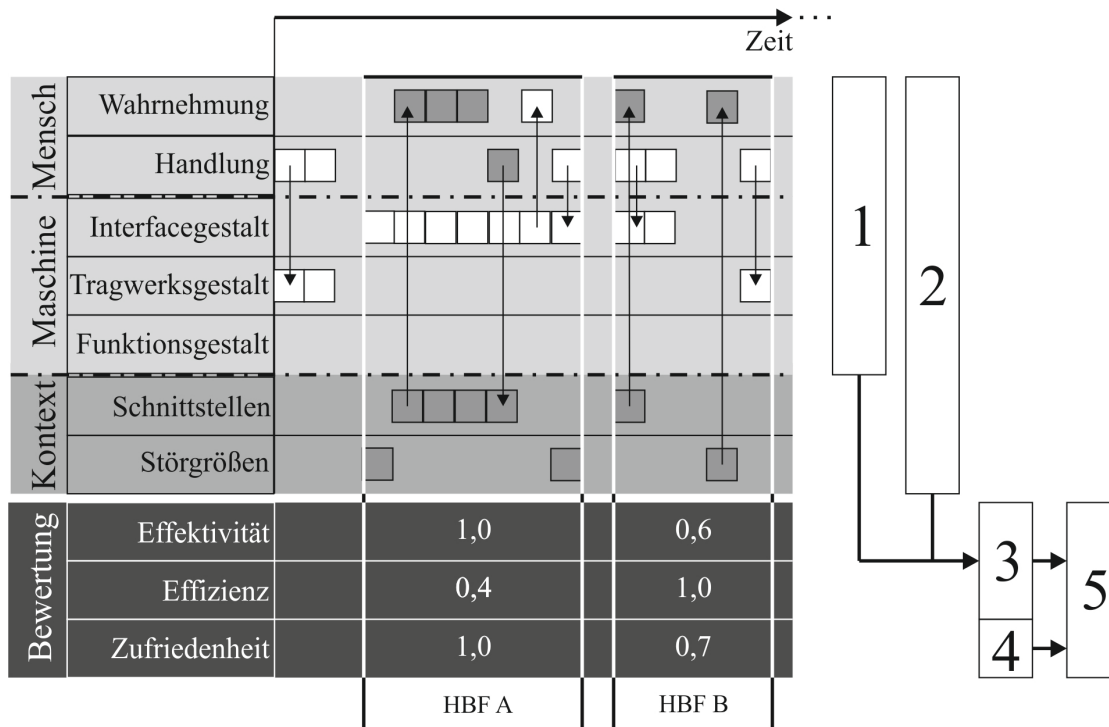


Abb.3: Ablauf einer Validierung mit der iFlow Methode

Der zweite Schritt des Auswertungsprozesses (vgl. Abb.3, Schritt 4) kann nur in Kooperation mit dem Bediener erfolgen. Hierzu ist eine post-observative Video-konfrontation notwendig. Nicht zuletzt werden hierbei auch Fehlzuschreibungen aus der Transkription beseitigt. Die subjektiv bewertete Zufriedenheit des Bedieners ergibt sich aus seiner Bewertung der Skalen der Benutzungsqualität des User Experience Questionnaire (Laugwitz et al. 2008). In einem finalen Bewertungsschritt, der wiederum einer Expertenevaluation bedarf, wird aus den drei generierten Usabilitymaßen ein Gesamtwert für jede Hauptbedienfunktion generiert (vgl. Abb.3, Schritt 5).

Exemplarische Validierung mit iFlow

Nachfolgend wird am Beispiel einer Spritzenpumpe die Vorgehensweise bei der Validierung mit iFlow nochmals verdeutlicht. Spritzenpumpen werden u.a. auf Intensivstationen eingesetzt. Hierbei ist zu beachten, dass diese selten alleinstehend, sondern immer in einem Spritzenpumpensystem, der sog. Infusionsampel angeordnet sind. Bei einem kritischen Zwischenfall kann es zu Stresssituationen kommen, welche Auswirkungen auf die Bedienung der Spritzenpumpe haben können. Ziel der Validierung ist der Nachweis der Gebrauchstauglichkeit auch in solchen (simulierten) Stresssituationen, bei denen mehrere Personen vielen verschiedenen und zudem verschachtelten Aufgaben nachgehen.

Im Rahmen des gebrauchstauglichkeitsorientierten Entwicklungsprozesses sind für die Spritzenpumpe acht Hauptbedienfunktionen (HBF) festgelegt worden: HBF A: Einschalten der Spritzenpumpe; HBF B: Spritze einlegen und sichern; HBF C: Spritze entlüften; HBF D: Manuellen Bolus injizieren; HBF E: Alarmsituation erkennen und beseitigen; HBF F: Laufrate eingeben; HBF G: Infusion starten; HBF H: Spritzenzustand überprüfen.

Um die Akzeptanzkriterien für die Hauptbedienfunktionen zu erhalten, wurden im Vorfeld der eigentlichen Validierung die expertenbasierten Bedienanalysen durchgeführt, video- und audiographiert und transkribiert. Dabei wurden Systemreaktionszeiten abgewartet, so dass ein nahezu idealer Bedienablauf aufgezeichnet wurde. Für das Usabilitymaß Effizienz wurde die Bediendauer und für das Usabilitymaß der Effektivität die notwendigen Bedienschritte sowie deren Permutation festgehalten.

Anschließend erfolgte die Durchführung der Simulation der Gerätebedienung im quasi-realen Kontext. Hierfür wurde das Einsatzgebiet der Intensivstation exemplarisch ausgewählt. Die Analysen fanden in vollausgestatteten Räumen eines deutschen Krankenhauses statt und wurden in Kooperation mit dem Tübinger Patientensicherheits- und Simulationszentrum durchgeführt, welches über moderne Full-Scale-Simulatoren verfügt. Es wurden acht Testdurchgänge durchgeführt von je 10 bis 15 Minuten Dauer. Vier Simulationen fanden vormittags und vier nachmittags statt. Insgesamt wurden drei verschiedene Szenarien simuliert, welche Notfallsituationen auf der Intensivstation darstellen, die auch für medizinisch erfahrenes Personal nicht zur alltäglichen Routine gehören (je drei Simulationen zu Szenario 1 und Szenario 3 und zwei Simulationen zu Szenario 2). Alle Testteilnehmer waren geübt in der Bedienung der Spritzenpumpe. Mehrere baugleiche Spritzenpumpen waren in einer Infusionsampel in einer dem Krankenhaus geläufigen makroergonomischen Anordnung und Raumposition platziert. Während der Simulation befanden sich nur die Akteure im Raum, deren Interaktionen von 2 Raumkameras aufgezeichnet wurden (makroskopisch). Des Weiteren wurde eine Kamera auf die Infusionsampel gerichtet (mikroskopisch).

Im Gegensatz zu Verifizierungsprozessen bzw. den damit verbundenen formativen Methoden, ist bei dieser Form der Validierung jeder Akteur in der Simulation potentieller Proband. Um möglichst realitätsnahe Situationen zu schaffen, werden keine Vorgaben bzgl. der Gerätebedienung gemacht. Für alle Sequenzen bei denen Interaktionen mit der Spritzenpumpe stattgefunden haben, werden aus den Videodaten iFlow-Diagramme erstellt und zusätzlich sowohl der Informationsaustausch mit den Schnittstellen des Kontexts als auch potentielle Störgrößen transkribiert. Anschließend werden die Hauptbedienfunktionen im iFlow-Diagramm markiert. Abbildung 4 zeigt exemplarisch die Visualisierung und Bewertung eines vollständig transkribierten iFlow-Diagramms eines Bedieners, welcher während eines Szenarios die Hauptbedienfunktionen A, B, F, G und H durchführte.

Validierung der Gebrauchstauglichkeit medizin- Technischer Produkte mit Hilfe der summativen Methode iFlow

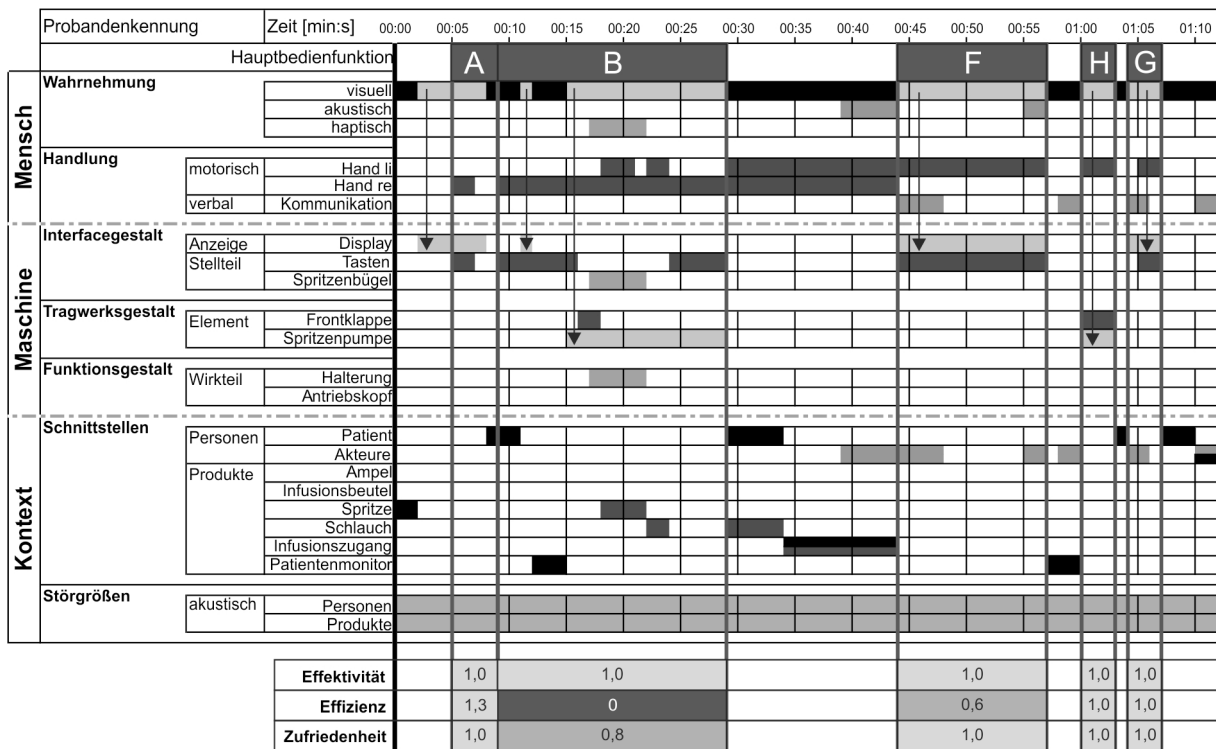


Abb.4: Visualisierung und Bewertung der Hauptbedienfunktionen im iFlow-Diagramm

Zu sehen ist, dass die Bedienszenarien überlagert sind mit stetigen akustischen Störgrößen von anderen Personen und Produkten. Um zu visualisieren, wie stark die Ablenkung von der eigentlichen Bedienaufgabe in den einzelnen Wahrnehmungskanälen ist, können die kontextbezogenen Interaktionen geschwärzt werden. In Abbildung 4 wurde dies für den visuellen Kanal durchgeführt.

Nachdem die Hauptbedienfunktionen zugeordnet und mit dem Bediener verifiziert wurden, erfolgt die Berechnung der Effektivitäts- und Effizienzwerte. Für Hauptbedienfunktion B in Abbildung 4 wurde bspw. die doppelte Bediendauer (20s Ist-Zeit zu 10s Soll-Zeit bedeutet eine Effizienzwertminderung um 100%) benötigt, was in dem minimalen Effizienzwert 0 resultiert. Dies liegt daran, dass eine Ablenkung durch den Patientenmonitor die Hauptaufgabe unterbrochen hatte und das Einlegen der Spritze in diesem Beispiel mehr Zeit in Anspruch genommen hat. Bei Hauptbedienfunktion F in Abbildung 4 wurde die Hauptaufgabe durch relevante Kommunikation mit einem anderen Akteur unterbrochen, weshalb nur ein Effizienzwert von $0,56=0,6$ erreicht werden konnte (13s Ist-Zeit zu 9s Soll-Zeit bedeutet eine Effizienzwertminderung um 44%). Da keine Bedienschritte ausgelassen bzw. fehlerhaft waren, wurde die Effektivität aller Hauptbedienfunktionen mit 1,0 bewertet. In der anschließenden Videokonfrontation mit dem Bediener wurden die Werte der Zufriedenheit aufgenommen. Der Wert entspricht dem gerundeten Mittel aus der Bewertung der Einzelskalen der Benutzungsqualität des User Experience Questionnaire (UEQ). In einem finalen, hier nicht weiter dokumentierten Schritt, mussten die drei Usabilitymaße für jede Hauptbedienfunktion miteinander in Beziehung gesetzt und zu einem Gesamtwert zusammengesetzt werden. Dieser Schritt wurde im Zuge der Ergebnisinterpretation wiederum expertenbasiert durch-

geführt. Die Kategorisierung der Gesamtwerte in sehr gute, gute und nicht befriedigende Ergebnisse wurde dann analog der VDI Richtlinie 2225 Blatt 3 vorgenommen (VDI 1998).

Diskussion

Der Aufwand der durch den Einsatz der iFlow Methode generiert wird, lässt sich nur bei der Validierung in komplexen Mensch-Maschine-Systemen rechtfertigen (vgl. Abb.2). Die Transkription stellt dabei noch den derzeit intensivsten Arbeitsschritt dar. Diese Umstände sollen aber durch ein teilweise automatisiertes Verfahren zukünftig verbessert werden. Der größte Nutzen liegt in der systematischen Quantifizierung der Gebrauchstauglichkeit der einzelnen Hauptbedienfunktionen und die damit verbundene Dokumentationsmöglichkeit. Auch hat sich herausgestellt, dass die Visualisierung über das iFlow Diagramm sich sehr gut eignet für die Diskussion in Entwicklungsteams oder Fokusgruppen. Die hier beschriebene exemplarische Analyse stellt streng genommen jedoch nur eine Validierung für Intensivstationen in dem entsprechenden Krankenhaus dar, da in anderen Einsatzgebieten oder Krankenhäusern die makro- und mikroergonomischen Voraussetzungen wieder ganz andere sein können. Nichtsdestotrotz konnten auch hier schon interessante zusätzliche Bedienaspekte beobachtet werden, welche die Gebrauchstauglichkeit beeinflusst haben wie bspw. die Erfüllung einer BediENAufgabe durch bis zu drei verschiedene Bediener, ungewöhnliche Bedienhaltungen oder die Bedienung bei Blendung oder Dunkelheit. Alle diese Aspekte werden dann ersichtlich, wenn das Bestreben besteht, die Validierung in nahezu realer Umgebung auch in extremen Situationen durchzuführen.

Literatur

- DIN Deutsches Institut für Normung e.V. (2011). Gebrauchstauglichkeit von Software 1 - Grundsätzliche Empfehlungen für Produkt- und Prozessgestaltung. Berlin: Beuth Verlag GmbH.
- Laugwitz, B.; Held, T. & Schrepp, M. (2008). *Construction and evaluation of a user experience questionnaire*. In: Holzinger, A. (Hrsg.): USAB 2008, LNCS 5298, 63-76.
- Morales, K. L.; Röbig, S. & Bruder, R. (2012): Learning from doing: Chances and constraints of studying medical devices through usability methods in field studies. *Zeitschrift für Arbeitswissenschaft*, 66, 115-127. Stuttgart: Ergonomia Verlag.
- Sarodnick, F. & Brau, H. (2011). Methoden der Usability Evaluation - Wissenschaftliche Grundlagen und praktische Anwendung. Bern: Hans Huber Verlag.

VDI-Richtlinie 2225, Blatt 3 (1998). Konstruktionsmethodik, Technisch-wirtschaftliches Konstruieren, Technisch-wirtschaftliche Bewertung. Düsseldorf: VDI Verlag.

Expertenbefragung: Erfassung der situationsbezogenen Wichtigkeit von Informationsquellen am Towerlotsenarbeitsplatz

Peer Manske¹, Max v. Schlippe², Michael Lange³
und Christoph Möhlenbrink¹

Schlüsselwörter: Visuelle Aufmerksamkeit, Fragebogen, SEEV, Towerlotsen

Kurzfassung

Towerlotsen sorgen dafür, dass der Verkehr an Flughäfen vor allem sicher, aber auch effizient und umweltschonend abläuft. Um ihre Aufgaben möglichst gut erfüllen zu können, bedienen sich Lotsen dazu visueller Informationen aus unterschiedlichen Quellen. Je nach Situation benötigen Lotsen unterschiedliche Informationen. Mit dem SEEV-Modell (z.B. Wickens & McCarley, 2007) kann die offene visuelle Aufmerksamkeit von Towerlotsen vorhergesagt werden. In dieser Studie steht besonders die Wichtigkeit von Informationsquellen im Vordergrund. Bei Lotsen, die sich durch eine hohe Expertise auszeichnen, wird davon ausgegangen, dass die Wichtigkeit einer Information auch den Fokus ihrer offenen visuellen Aufmerksamkeit stark beeinflusst. Um fundierte Aussagen über die Wichtigkeit von Areas of Interest zu treffen, konnten Werte durch die Einschätzung mehrerer Experten ermittelt werden. Zu diesem Zweck wurde ein Fragebogen entwickelt und eine Befragung von Towerlotsen durchgeführt, deren Ergebnisse hier berichtet werden. Die ermittelten Wichtigkeiten können zur Erstellung eines Wahrscheinlichkeitsmodells der offenen visuellen Aufmerksamkeit von Lotsen verwendet werden.

Aufgabe von Towerlotsen

Towerlotsen sind an Flughäfen dafür verantwortlich, dass der Verkehr (besonders der Flugverkehr) sicher und effizient abgewickelt wird. Sie stehen mit allen Luftfahrzeugen (Lfz), Fahrzeugen und Personen in ihrem Zuständigkeitsbereich in Kontakt und koordinieren sie mittels Anweisungen und Freigaben direkt per Sprechfunk. Neben dem sprachlichen Informationsaustausch gibt es weitere Informationen, die ein Lotse verwendet, um ein aktuelles Bild der Situation aufrecht zu erhalten (Pinska & Bourgois, 2007). Diese Informationen werden zumeist visuell von Anzeigen und anderen Instrumenten sowie aus der Sicht nach draußen abgerufen. Abb.13 stellt schematisch die wichtigsten visuellen Informationsquellen dar.

¹ Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Lilienthalplatz 7, 38108 Braunschweig

² Technische Universität Berlin, Marchstrasse 23, 10587 Berlin

³ The Tower Company GmbH, Ohmstraße 12, 63225 Langen

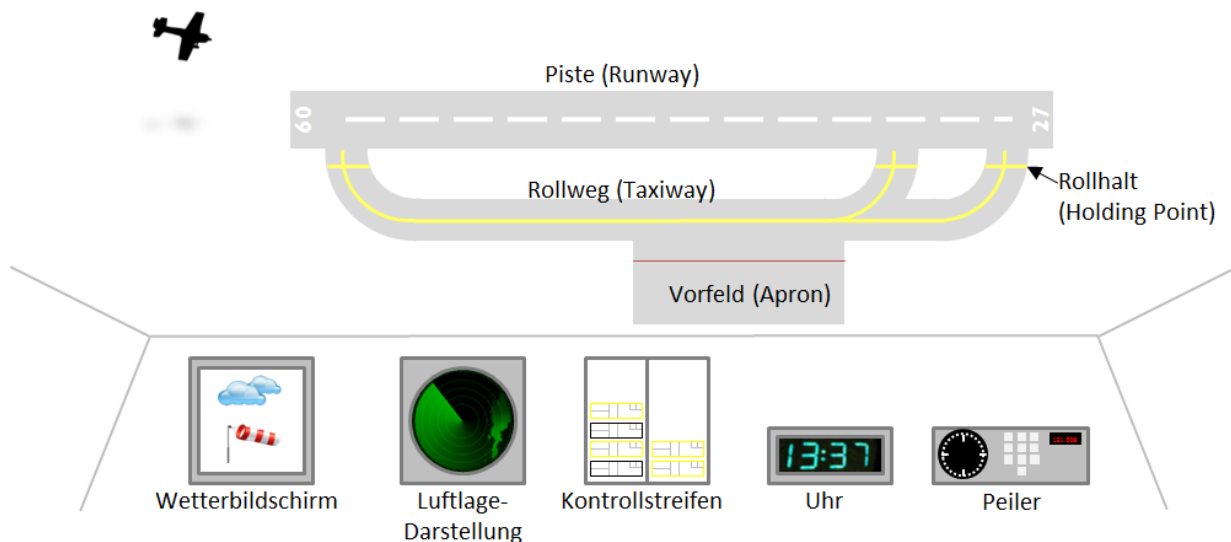


Abb.13: Schematische Darstellung eines Lotsenarbeitsplatzes

Die Außensicht (oberhalb der Instrumente dargestellt) gewährt dem Lotsen die Übersicht über die kontrollierten Betriebsflächen des Flughafens und dessen umgebenden Luftraum. Die Instrumente im Tower dienen als zusätzliche Hilfen. Der Wetterbildschirm zeigt unter anderem den aktuellen Wind und Luftdruck an. Diese Informationen muss der Lotse den Piloten startender und landender Lfz mitteilen. Die Luftlagedarstellung dient der Positionsinformation und Ortung von Lfz. Der Lotse kann mit ihrer Hilfe ein gutes Bild der Situationen behalten und besser planen. Kontrollstreifen helfen dem Lotsen bei der Planung und Koordination des Luftverkehrs, und fungieren als eine Art Gedächtnisstütze. Dabei werden die wichtigsten Informationen und Anweisungen für jeden Flug festgehalten. Die exakte Uhrzeit spielt bei der Kontrolle und Koordination des Luftverkehrs eine entscheidende Rolle. So dürfen z.B. bestimmte Lfz nur in einem bestimmten Zeitfenster starten. Um zeitbedingte Koordinationsfehler zu minimieren, wird in der Flugsicherung weltweit mit der koordinierten Weltzeit (UTC) gearbeitet. Mit Hilfe des Peilers kann der Lotse sehen, aus welcher Richtung ein Funkspruch abgegeben wird und einem Lfz zuordnen. Die hier nur kurz beschriebenen Instrumente sind lediglich eine Auswahl der Hilfsmittel im Tower. Es handelt sich bei dieser Auswahl um visuelle Informationsquellen, die als besonders wichtig für die Flugverkehrskontrolle an einem Tower identifiziert werden konnten (Schlippe & Manske, 2013). Dabei wurde sich an Gesprächen mit Towerlotsen, Towerbegehungen und an den Beschreibungen von Pinska & Bourgois (2007) orientiert.

Visuelle Aufmerksamkeit

Um die Aufgabe der Flugsicherung besser zu verstehen, ist es wichtig zu ermitteln, wie Lotsen Informationen visuell abrufen, um ein adäquates Bild der Situation zu behalten. In einigen Untersuchungen mit Towerlotsen wurde bereits ermittelt, wie die offene visuelle Aufmerksamkeit (z.B. McCarley & Kramer, 2008; Posner 1980) von Lotsen auf die verschiedenen Informationsquellen (Areas of Interest,

AOIs) generell verteilt ist (z.B. Pinska & Bourgois, 2007; Oehme & Schulz-Rueckert, 2010).

Flugsicherung ist allerdings sehr dynamisch, weil sie ständige Überprüfung und Anpassung an die aktuelle Lage erfordert. Jede Situation erfordert dabei Entscheidungen des Lotsen, für die er situationsspezifisch verschiedene Informationen berücksichtigen muss. Entsprechend unterschiedlicher Situationen, unterscheiden sich somit die Aufgaben des Lotsen und folglich sein Informationssuchverhalten. (Möhlenbrink, Papenfuß, Friedrich & Jipp, 2011). Es ist also zentral, nicht nur die gesamte Blickverteilung über die Zeit zu beachten, sondern auch die Wahrscheinlichkeit von Blicken bei unterschiedlichen Anforderungen der Situation zu ermitteln.

Vorhersage visueller Aufmerksamkeit

Um offene visuelle Aufmerksamkeit vorauszusagen, hat das SEEV-Modell (z.B. Wickens & McCarley, 2007) bereits häufig Anwendung gefunden. Es eignet sich dazu, die Aufmerksamkeitszuwendung von Menschen in komplexen Arbeitssituationen (Flug- und Fahraufgaben) zu modellieren (Wickens, Helleberg, Goh, Xu & Horrey, 2001; Wickens, McCarley & Thomas, 2003; Wickens & McCarley, 2007).

Das SEEV-Modell berücksichtigt vier Komponenten zur Vorhersage der Blick-Wahrscheinlichkeit: Salienz (Saliency), Aufwand (Effort), Erwartung (Expectancy) und Wichtigkeit (Value). Salienz und Aufwand werden dabei als Bottom-Up-Prozesse betrachtet und sind damit eher als Eigenschaft der AOI zu verstehen. Erwartung und Wichtigkeit werden den Top-Down-Prozessen zugeordnet und erfordern somit willentliche Aufmerksamkeitszuwendung. Die vier Komponenten bilden ein analytisches und statisches Modell, welches durch die Formel „ $P(AOI) = S - Ef + Ex + V$ “ angegeben werden kann und – über mehrere AOIs gerechnet – eine generelle Verteilung der Aufmerksamkeit ergibt. Eine detaillierte Beschreibung des Modells findet sich u.a. bei Wickens et al. (2007).

Forschungsfokus

Um das SEEV-Modell auf spezifische Situationen der Flugsicherung anzuwenden, wurde in dieser Untersuchung der Komponente Wichtigkeit (Value) eine bedeutende Rolle zugeschrieben. Diese Komponente beschreibt, wie wichtig eine AOI für eine bestimmte Aufgabe ist. Es wird angenommen, dass je nach Aufgabe naturgemäß verschiedene Informationen aus unterschiedlichen Informationsquellen zur Entscheidungsfindung benötigt und daher auch vom Lotsen fixiert werden. Andere Bereiche sind situationsabhängig hingegen weniger wichtig, wodurch die Wahrscheinlichkeit, diese zu betrachten, auch geringer ausfallen sollte. Die anderen Komponenten erscheinen nachrangig, da die Erwartung, dass sich etwas verändert, in kurzen, klar definierten Situationen eher marginal ist und Bottom-Up-Prozesse (Salienz und Aufwand) bei Profis in Standardsituationen ebenfalls ver-

nachlässigt werden können (Wickens, Goh, Helleberg, Horrey & Talleur, 2003, S.373).

Die letzte Komponente des SEEV-Modells ist in dieser Arbeit zentral. Dabei stellt sich die Frage, auf welche Art und Weise bei SEEV konkrete Werte für die Wichtigkeit je Situation zugeordnet werden können. Gemäß Wickens et al. (2001) werden die vergebenen Werte der Literatur oder dem allgemeinen Verständnis entnommen. Die Wichtigkeit einzelner AOIs für bestimmte Aufgaben erhalte man durch eine kognitive Aufgabenanalyse (Wickens et al. 2003). Für eine vertiefende Betrachtung der Wichtigkeit ist es sinnvoll, die Wichtigkeit situationsspezifisch zu bewerten, da sich Aufgaben in verschiedenen Situationen unterscheiden (vgl. Moehlenbrink, Papenfuß, Friedrich & Jipp, 2011). Aus diesem Grund wurde ein Fragebogen entwickelt, um Towerlotsen als Experten direkt zu ihrer Einschätzung der Wichtigkeit ihrer Arbeitsmittel in definierten Situationen zu befragen. Die so ermittelten Werte der Wichtigkeit sollen als Grundlage für die Berechnung situationsspezifischer visueller Aufmerksamkeit dienen.

Stichprobe

In einer Studie von Lange (im Druck) wurden Towerlotsen der The Tower Company GmbH als Experten mit dem von Schlippe und Manske (2013) entwickelten Fragebogen befragt. In die Analyse gingen die Daten aus insgesamt 18 Fragebögen ein. Die Towerlotsen im Alter zwischen 23 und 43 Jahren ($M_{(Alter)} = 33,1$ Jahre) hatten eine durchschnittliche Berufserfahrung von $M_{(Erfahrung)} = 9,6$ Jahren.

Fragebogenerhebung

Von Schlippe und Manske (2013) wurde ein Fragebogen zur Einschätzung der Wichtigkeit von AOIs bei Towerlotsen entwickelt. Dabei wurde zunächst eine schematische Darstellung eines Lotsenarbeitsplatzes erstellt (vgl. Abb.13). Bei der Betrachtung der AOIs war zudem wichtig, dass die Außensicht nicht als einzelne AOI betrachtet wird, sondern aus vielen Unterteilungen bestehen kann (z.B. Anflugbereich, Piste etc.).

Anschließend wurden einfache Standardsituationen identifiziert, wie sie an Flugplätzen (hier solche mit nur einer Piste) auftreten können und Entscheidungen des Lotsen erfordern. Solche Situationen sind Starts und Landungen mit ihren verschiedenen Phasen, sowie das Rollen am Boden. Auch Kombinationen der einfachen Situationen mit maximal zwei Lfz wurden berücksichtigt.

Um die Lotsen zu den Situationen zu befragen, wurde für jede Situation ein einzelner Fragebogen erstellt, in dem diese in der schematischen Außensicht bildlich dargestellt und zusätzlich kurz beschrieben wurde.

Die Aufgabe der Lotsen war im ersten Schritt, die relevanten AOIs zu markieren. Sie sollten eine Auswahl treffen, welche Informationsquellen in der beschriebenen Situation wichtig sind. Danach sollten die Lotsen bewerten, für wie wichtig sie die

gewählten AOIs in der jeweiligen Situation einschätzten und sie in eine Rangfolge bringen.

Zur Interpretation der Wichtigkeit einzelner AOIs wurden anschließend die Einschätzungen aller Lotsen pro Situation gemittelt und als relative Werte dargestellt.

Ergebnisse

In der Analyse der Lotseneinschätzungen wurden ordinale Werte errechnet, welche die relative Wichtigkeit der AOIs untereinander in den einzelnen Situationen darstellen und sich zur einfachen Vergleichbarkeit in jeder Situation zum Wert 1 aufsummieren. Diese sind in Tab.2 zusammengefasst. In Situationen, in denen ein Lfz nicht vorhanden war oder nicht als sichtbar angenommen wurde, wurden keine Werte vergeben.

Tab.2: Relative eingeschätzte Wichtigkeit der AOIs pro Situation

Situation	Lfz A	Lfz B	Anflug	RWY	Rollhalt	Abroller	TWY	Vorfeld	Abflug	Wetter	Luftlage	Kontrollstreifen	Uhr	Peiler
Anflug auf ILS	/	/	.18	.14	.00	.00	.00	.00	.02	.14	.22	.20	.03	.08
Landefreigabe	.14	/	.18	.21	.01	.01	.00	.00	.01	.17	.08	.15	.04	.01
Landung	.18	/	.00	.10	.02	.20	.05	.10	.00	.01	.02	.19	.12	.00
Lfz erbittet Rollfreigabe	.15	/	.00	.00	.03	.01	.22	.16	.00	.10	.02	.19	.10	.03
Startfreigabe	.08	/	.07	.20	.00	.01	.00	.00	.08	.15	.16	.16	.09	.00
Freigabe zum Aufrollen	.09	/	.11	.18	.08	.01	.00	.00	.06	.11	.14	.15	.07	.01
Lfz ist abgehoben	.23	/	.00	.03	.00	.00	.00	.00	.24	.02	.23	.14	.11	.00
Lfz A im Anflug, Lfz B ist gelandet und noch auf der Piste	.14	.14	.16	.13	.00	.11	.00	.02	.00	.11	.04	.09	.04	.00
Lfz A ist im Anflug, Lfz B rollt zum Rollhalt	.15	.14	.15	.15	.14	.02	.01	.00	.01	.08	.06	.08	.02	.00
Lfz A steht am Rollhalt, Lfz B ist gelandet und noch auf der Piste	.11	.15	.06	.11	.10	.10	.01	.01	.04	.07	.10	.12	.05	.00
Lfz A steht am Rollhalt, Lfz B ist gerade gestartet	.09	.14	.07	.11	.08	.01	.00	.00	.13	.08	.12	.12	.05	.01

Aus der Tabelle lassen sich die einzelnen Werte pro AOI und Situation ablesen. So lässt sich beispielsweise erkennen, dass bei Lande- und Startfreigabe der Piste mit relativen Werten von $W_{(AOI)} \geq .20$ eine große Wichtigkeit beigemessen wird. Bei einer Startfreigabe wird zusätzlich die Luftlagedarstellung als eher relevant

bewertet ($W_{(AOI)} = .16$), wohingegen sie bei der Landung ($W_{(AOI)} = .08$) weniger wichtig erscheint. Eine Erklärung könnte hier die Überprüfung des Abflugsektors (der Luftraum, in den das Lfz nach dem Abheben einfliegt) mittels Luftlagedarstellung durch den Lotsen bieten. Bei einer Landung wird dieser Luftraum im Regelfall nicht beflogen (außer, ein Lfz bricht den Landeanflug ab und startet durch), und da das anfliegende Lfz nachfolgend auf der Piste landen soll, erscheint eine Betrachtung dieser eine höhere Wichtigkeit zuzukommen, als der Luftlagedarstellung zur Beobachtung des umgebenden Luftraums.

Um den Umfang dieser Studie nicht zu überschreiten, soll an dieser Stelle lediglich ein exemplarisches Ergebnis präsentiert werden. Weitere Ergebnisse sind auf Anfrage bei den Autoren erhältlich.

Diskussion und Ausblick

Die Fragebogenanalyse eignet sich, um Einschätzungen zur situationsabhängigen Wichtigkeit von Informationsquellen zu erhalten, welche aus reinen Blickdaten nicht extrahiert werden können. Auch erscheint die situationsspezifische Befragung als geeignete Methode, um detaillierte Werte der Wichtigkeit für verschiedene Aufgaben des Towerlotsen zu ermitteln. Die gewonnenen Daten bieten zukünftig die Möglichkeit, die Diversität der Einschätzungen mehrerer Lotsen zu berücksichtigen, um abweichendes Blickverhalten erklärbar zu machen. Zudem ist eine Analyse der numerischen Wichtigkeitseinschätzung möglich, die den AOIs detailliertere Werte als Rangfolgewerte zuweist.

Ob die mittels Fragebogen erhobenen deskriptiven Einschätzungen der Wichtigkeit tatsächliches Blickverhalten abbilden, bleibt in einer anschließenden Validierung zu klären. Dabei sollte ein Hauptaugenmerk nicht nur auf den Übereinstimmungen zwischen eingeschätzten und tatsächlichen Werten liegen, sondern besonders auch auf den Abweichungen zur Vorhersage durch die deskriptiven Daten. So könnten Hypothesen erarbeitet werden bezüglich weiterer Faktoren, welche die Blickverteilung von Towerlotsen beeinflussen.

Literatur

- Lange, M. (im Druck). Erfassung visueller Informationsquellen von Towerlotsen. Braunschweig: Deutsches Zentrum für Luft- und Raumfahrt.
- McCarley, J. S., & Kramer, A. F. (2008). Eye Movements as a Window on Perception and Cognition. In R. Parasuraman & M. Rizzo (Eds.), *Neuroergonomics: The Brain at Work* (pp. 95-112). New York, NY, USA: Oxford University Press, Inc.
- Moehlenbrink, C., Papenfuß, A., Friedrich, M., & Jipp, M. (2011). Monitoring behaviour of tower controllers. In D. de Waard, N. Gérard, L. Onnasch, R. Wiczorek, & Dietrich Manzey (Eds.), *Human Centred Automation* (pp. 269-283) Maastricht, the Netherlands: Shaker Publishing.

- Oehme, A., & Schulz-Rueckert, D. (2010). Distant Air Traffic Control for Regional Airports. *IFAC Symposium on Analysis, Design and Evaluation of Human-Machine Systems*. Valenciennes, Frankreich.
- Pinska, E., & Bourgois, M. (2007). Warsaw Tower Observations. *Eurocontrol Experimental Centre, Report 2007*. Bretigny-sur-Orge, France.
- Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32, 3-25.
- Schlippe, M., & Manske, P. (2013). Entwicklung eines Fragebogens zur Erfassung der Wichtigkeit visueller Informationsquellen bei Towerlotsen. *Institutsbericht No. 112-2013/30*, Braunschweig: Deutsches Zentrum für Luft- und Raumfahrt.
- Wickens, C.D., Goh, J., Helleberg, J., Horrey, W.J., & Talleur, D.A. (2003). Attentional Models of Multitask Pilot Performance Using Advanced Display Technology. *Human Factors*, 3, 360-380.
- Wickens, C.D., Helleberg, J., Goh, J., Xu, X., & Horrey, W.J. (2001). Pilot Task Management: Testing an Attentional Expected Value Model of Visual Scanning (ARL-01-14/NASA-01-7). Savoy, IL: University of Illinois, Aviation Research Lab.
- Wickens, C. D., & McCarley, J. S. (2007). *Applied Attention Theory*. Boca Raton, FL: CRC Press.
- Wickens, C.D., McCarley, J.S., Alexander, A.L., Thomas, L.C., Ambinder, M., & Zheng, S. (2007). Attention-Situation Awareness (A-SA) Model of Pilot Error. In D.C. Foyle & B.L. Hooey (Eds.), *Human Performance Modeling in Aviation*. Mahwah, NJ: Lawrence Erlbaum.
- Wickens, C. D., McCarley, J., & Thomas, L. (2003). Attention-Situation Awareness (A-SA) Model. In D.C. Foyle, A. Goodman & B.L. Hooey (Eds.), *Proceedings of the 2003 Conference on Human Performance Modeling of Approach and Landing with Augmented Displays* (pp. 189-225). Moffett Field, CA: NASA.

Entwicklung eines generischen Flughafen-Modells für die effizientere Makrokognitive Modellierung des Mensch-Maschine-Systems der Flughafenverkehrskontrolle mit farbigen Petri-Netzen

Peer Manske¹, Hardy Smieszek², Andreas Hasselberg¹
und Christoph Möhlenbrink¹

Schlüsselwörter: Petri-Netze, Modellierung, Flughafenprozesse, Makrokognition

Zusammenfassung

Die Verwendung kognitiver Modelle zur Analyse menschlichen Verhaltens und menschlicher Kognition in soziotechnischen Systemen stellt eine vielversprechende Methode der Mensch-Maschine-Forschung dar. Zur Untersuchung der Tätigkeiten von Towerlotsen werden makrokognitive Modelle auf der Basis farbiger Petri-Netze verwendet. Da die Arbeit von Lotsen nicht losgelöst vom Flughafen und den darauf ablaufenden Prozessen betrachtet werden kann, muss zum Lotsen-Modell auch ein Flughafenmodell erstellt werden. Bislang erfolgte die Modellierung der operativen Bereiche eines Flughafens individuell und entsprechend der räumlichen Struktur der Vorlage. Zur Anpassung des Modells an einen anderen Flughafen musste daher stark in die Struktur eingegriffen werden, was sehr zeitaufwendig und anfällig für Fehler ist. Es wird ein Ansatz präsentiert, welcher es erlaubt mit Hilfe eines generischen Flughafenmodells sowohl die Effizienz der Erstellung zu erhöhen als auch die Fehleranfälligkeit zu reduzieren. Damit wird ein Mittel bereitgestellt, um ein Lotsen-Flughafen-Modell ohne strukturelle Veränderungen schnell und effizient an verschiedene Flughäfen anzupassen und zu simulieren.

Einleitung

Die Aufgabe von Towerlotsen ist die Koordination der Bewegung von (Luft-)Fahrzeugen in der Luft und am Boden. Die Menge und Komplexität des Verkehrs sowie die Komplexität des Flughafens bestimmen die Anforderungen an die Lotsen. Auch die Arbeitsweise und -beanspruchung der Lotsen hat Auswirkungen auf den kompletten Flughafenprozess. Um die Art und Weise ihrer Arbeit und ihre Entscheidungsprozesse besser zu verstehen, werden kognitive Modelle erstellt. Da die Arbeit von Lotsen nicht losgelöst vom Flughafen betrachtet werden kann, werden häufig makrokognitive Modelle auf Basis farbiger Petri-Netze als Ansatz gewählt, welche bestimmte Aspekte der Entscheidungsfindung und andere kognitive Komponenten von Lotsen und deren Einfluss auf das Gesamtsystem modellieren

¹ Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Lilienthalplatz 7, 38108 Braunschweig

² Technische Universität Berlin, Marchstrasse 23, 10587 Berlin

Entwicklung eines generischen Flughafen-
Modells für die effizientere Makrokognitive Modellierung des
Mensch-Maschine-Systems der Flughafenverkehrskontrolle mit farbigen Petri-Netzen

(z.B. Werther, 2006; Werther, Möhlenbrink & Rudolph, 2007; Smieszek, Manske, Hasselberg, Rußwinkel & Möhlenbrink, 2013).

Diese Modelle bestehen im Allgemeinen aus einem Lotsen-, einem Interaktions- und einem Flughafenmodell, welches die am Flughafen ablaufenden Prozesse beschreibt. Das Interaktionsmodell definiert den Informationsaustausch zwischen Lotsen- und Flughafenmodell (Cacciabue, 1998).

Der Anspruch an ein allgemein gültiges Lotsen-Modell ist, dass es nicht nur mit verschiedenen Verkehrssituationen umgehen, sondern auch mit unterschiedlichen Flugplätzen interagieren kann, die sich in ihrer Auslegung und Komplexität unterscheiden. Bislang erfolgte die Modellierung der operativen Bereiche eines Flughafens individuell und entsprechend der räumlichen Struktur der Vorlage. Zur Anpassung des Modells an einen anderen Flughafen musste daher stark in die Struktur eingegriffen werden. Solche Veränderungen sind sehr zeitaufwendig und anfällig für Fehler. Zudem können Änderungen im Flughafenmodell Änderungen im Interaktionsmodell nach sich ziehen. Der hier vorgestellte Ansatz eines generischen Flughafenmodells modelliert den Flughafen stattdessen entsprechend seiner funktionalen Struktur. Dabei bleibt die Prozesslogik für jeden Flughafen gleich. Die individuelle Struktur eines Flughafens kann hingegen variabel geladen werden.

Ziel ist zu zeigen, dass die Übertragung von Lotsen-Flughafen-Modellen auf unterschiedliche Flughäfen mit dem generischen Flughafenmodell stark vereinfacht werden kann.

Mensch-Maschine-Modellierung mit farbigen Petri-Netzen

Im Folgenden werden zunächst die Grundlagen des verwendeten Modellierungswerkzeugs, der farbigen Petri-Netze, erläutert. Diese werden auf das Wesentliche, für das Verständnis der folgenden Abschnitte notwendige, reduziert. Eine ausführliche Einführung in Petri-Netze kann beispielsweise bei Baumgarten (1996) oder Reisig (2010) gefunden werden, eine Einführung in farbige Petri-Netze im speziellen bei Jensen und Kristensen (2009).

Farbige Petri-Netze

Petri-Netze sind ein ereignisdiskretes Modellierungswerkzeug und ermöglichen eine graphische Programmierung. Ein Petri-Netz besteht generell aus zwei verschiedenen Gruppen von Elementen: Plätze und Transitionen. Plätze (als Ovale dargestellt) können über Kanten (gerichtete Pfeile) nur mit Transitionen (Rechtecke) verbunden werden und vice versa (vgl. Abbildung 14). Plätze sind passive Elemente und dienen dem Speichern von Informationen. Jede Information wird als Marke (auch Token genannt) auf den entsprechenden Platz abgelegt. Im Gegensatz dazu sind Transitionen aktive Elemente, die Marken auf direkt verbundenen Plätzen verändern können. Durch Schalten einer Transition werden Marken auf deren Vorplätzen (Kante zur Transition gerichtet) gelöscht und auf deren Nachplätzen

(Kante zum Platz ausgerichtet) erzeugt. Farbige Petri-Netze ermöglichen es im Gegensatz zu klassischen Petri-Netzen, Marken unterschiedlicher Datentypen und Eigenschaften zu verwenden, so dass eine Marke beispielsweise ein Flugzeug (mit Typ, Rufzeichen, Ziel, usw.) repräsentieren kann. Zur Erstellung von farbigen Petri-Netzen wird die Software *CPN Tools* (siehe Jensen & Kristensen 2009) verwendet.

Flughafenspezifische Modellierung

Wie oben beschrieben, kann die Analyse der Tätigkeiten von Lotsen mit Hilfe der makrokognitiven Modellierung mit Petri-Netzen erfolgen. Diese beinhaltet neben dem Lotsenmodell auch die Erstellung eines Flughafenmodells, mit dem das Lotsenmodell über das Interaktionsmodell interagiert. In der Vergangenheit wurde dazu meist ein bestimmter exemplarischer Flughafen modelliert (z.B. Moehlenbrink, Manske & Kirlik, 2012; Davidrajuh & Lin, 2011). Gemäß bisheriger Logik der Erstellung eines Flughafenmodells entsprach dabei die Struktur des Petri-Netzes auch dem räumlichen Aufbau der Betriebsflächen am Vorbild. Somit wurde die Bewegung eines Luftfahrzeuges auf dem Flughafen folgendermaßen abgebildet: Das Luftfahrzeug selbst wurde als Marke modelliert, jede Fläche des Flughafens hingegen als Platz. Eine entsprechende Marke bewegte sich über eine Transition von einem Platz zum nächsten Platz weiter, was beispielsweise der Bewegung von einem Abschnitt eines Rollweges zum nächsten entsprach. So ergab sich bei bisheriger Modellierung eine Vielzahl an Plätzen, deren Verbindungen untereinander starr festgelegt waren. Außerdem existierten von den verschiedenen Plätzen Verbindungen zum Interaktionsmodell, um den Status der Flugzeuge abzufragen und die Flugzeuge zu beeinflussen. Um das Lotsenmodell mit einem anderen Flughafen interagieren zu lassen, war die notwendige Anpassung daher mit einer aufwendigen Umstrukturierung verbunden. Das liegt daran, dass je nach Ausmaß der Umstrukturierung viele Elemente gelöscht oder neu erzeugt, sowie alle Elemente untereinander neu verbunden werden mussten und dabei an jedem Element erneut Fehler entstehen konnten.

Generisches Lotsen-Flughafen Modell

Invariable Prozesslogik

Das hier vorgestellte generische Petri-Netz-Modell eines Flughafens folgt der Idee, dass einige wenige Orte mit derselben Funktionalität an jedem Flughafen gleich sind; es gibt eine oder mehrere Pisten mit An- und Abflugbereichen, Rollwege und Stellplätze für Luft- und andere Fahrzeuge. Statt jede dieser Flächen einzeln zu modellieren, wurde für das generische Modell jeweils ein einziger Platz pro funktionalen Bereich angelegt (s. Abbildung 1). Die Stellplätze (Stands), Rollwege (Taxiways, TWYs), Pisten (Runways, RWYs) sowie Endanflug (Final) und Abflug (Departure) sind demnach im Petri-Netz durch nur jeweils einen Platz modelliert.

Die Übergänge von einem funktionalen Bereich zum nächsten sind über Transitionen beschrieben. Sie bilden zusammen die invariable Prozesslogik. Diese Grundstruktur ist dabei über das Interaktionsmodell fest mit dem Lotsenmodell verbunden und bleibt gänzlich unverändert, wenn die variable Prozessstruktur des Flughafenmodells geändert wird.

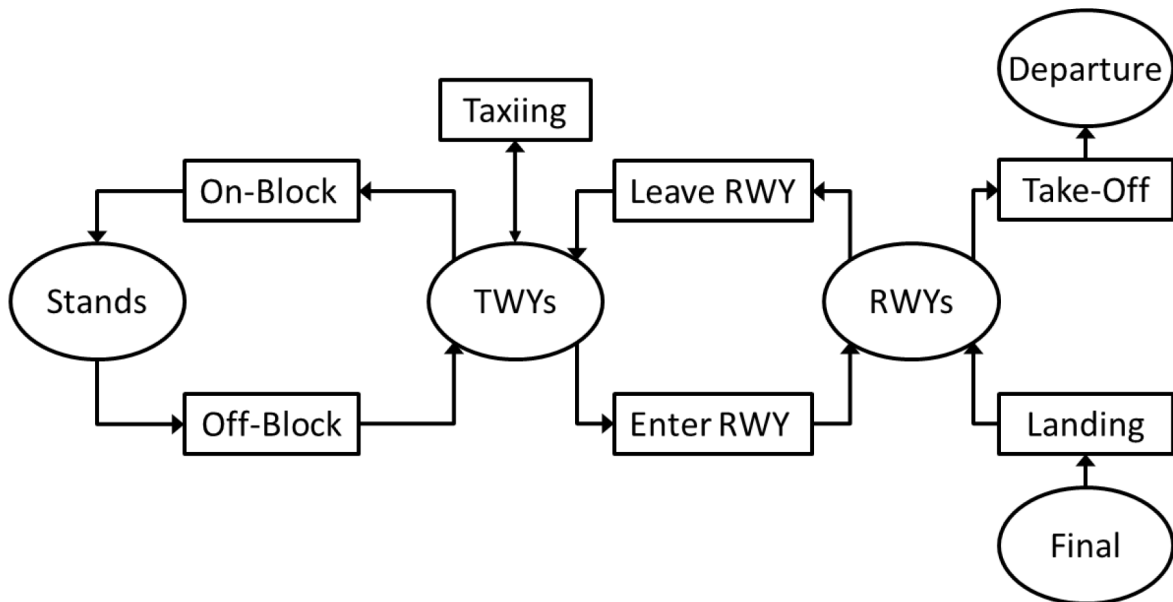


Abbildung 14: Elemente der invariablen Prozesslogik

Variable Prozessstruktur

Flughäfen unterscheiden sich untereinander teils stark hinsichtlich ihres Aussehens und ihrer Komplexität. Der neue Hauptstadtflughafen Berlin Brandenburg „Willy Brandt“ (im Bau) beispielsweise wird mit seinen zwei Pisten nördlich und südlich des Flughafengebäudes bereits deutlich komplexer sein als der Flughafen Braunschweig-Wolfsburg mit nur einer Piste. Trotzdem ist der Anspruch an ein Lotsenmodell, den Verkehr an beiden Flughäfen abarbeiten zu können. Die unterschiedlichen Informationen über die Struktur der einzelnen Elemente und ihrer Verbindungen untereinander am Flughafen werden im generischen Flughafenmodell separat und unabhängig von der invariablen Prozesslogik abgelegt. Sie werden in Form von Marken in das Grundgerüst der invariablen Prozesslogik geladen. Die Marken werden wiederum in Form von Variablen dazu in externen Textdateien erstellt und später nur noch in das Modell in CPN Tools eingelesen.

Die Informationen bilden die variable Prozessstruktur. Sie beinhaltet

- a. eine Definition aller Stellplätze, Rollwegabschnitte, Pisten, An- und Abflüge sowie deren Bezeichnungen (Location),
- b. die Verbindungen der einzelnen Abschnitte von Rollwegen miteinander sowie die Anknüpfung an Stellplätze, andere Rollwege oder Pisten (Connections)

- c. und die Zeitdauern, welche Flugzeuge benötigen, um diese Abschnitte zu passieren (Duration).

Daraus gestaltet sich die funktionale Basis eines Flughafens mit allen Spezifikationen. Um auf dieser Basis einen Flughafen dynamisch zu simulieren, können in einer zusätzlichen austauschbaren Szenario-Datei Objekte (Vehikel) hinzugefügt werden, die sich über die Betriebsflächen bewegen. Jedes Fahrzeug wird durch eine Marke repräsentiert die unter anderem die aktuelle Position des Fahrzeugs enthält. Bewegt sich das Fahrzeug weiter, wird diese Information aktualisiert und der nächste Abschnitt eingetragen. Darüber hinaus wird die Marke der verlassenen Fläche (Locations) als frei markiert, während die nächste Fläche als belegt markiert wird. Bewegungen über die Betriebsflächen erfolgen anhand der hinterlegten Verbindungen (Connections) und dauern entsprechend der definierten Passierzeit an (Duration).

Beim Einlesen der Textdateien erfolgt ein Syntaxcheck, sodass Syntaxfehler direkt erkannt und behoben werden können. Die so in CPN Tools eingebundenen Dateien definieren dann die anfängliche Markenbelegung des Flughafenmodells.

Lotsen- und Interaktionsmodell

Das Lotsen- sowie das Interaktionsmodell sollen allgemeingültig sein, also mit unterschiedlichen Flughafenlayouts umgehen können. Um an verschiedenen Flughäfen arbeiten zu können, ist ein unterschiedliches Hintergrundwissen über die Betriebsprozeduren erforderlich, die die möglichen Freigaben festlegen. Wenn im generischen Modell ein anderer Flughafen geladen wird, so muss auch eine für diesen Flughafen definierte Logik-Datei mit Betriebsprozeduren geladen werden, die als prozedurale Wissensbasis für das Lotsenmodell dient. Dem Lotsenmodell wird auf diesem Wege Expertenwissen darüber vermittelt, welche Rollrouten möglich sind und welche Freigaben somit erteilt werden können.

Ergebnisse

Mit Hilfe des generischen Flughafenmodells wurde exemplarisch der Aufbau zweier Flughäfen umgesetzt. Zunächst entstand für eine erste Simulation der Flughafen Braunschweig-Wolfsburg (EDVE) mit nur einer Start-/Landebahn. Um die Stärke des generischen Ansatzes zu zeigen, wurde noch die Struktur des sehr viel komplexeren Flughafens Berlin-Brandenburg-International (EDDB) implementiert. Verglichen mit der bisher üblichen Umsetzung von Flughafenmodellen in CPN Tools, erweist sich die Beschreibung der variablen Prozessstruktur in Textdateien als deutlich effektiver und innerhalb von wenigen Stunden bis maximal 1-2 Werktagen (je nach Komplexität des Flughafens) sehr zügig zu erledigen. Dagegen kann die Neuerstellung bzw. Abänderung der Netzstruktur eines Flughafenmodells im herkömmlichen Sinne je nach Komplexität des jeweiligen Flughafens mehrere Tage bis Wochen in Anspruch nehmen. Hierbei können an jeder bearbeiteten Verbindung Fehler entstehen, welche zu beheben wiederum Zeit kostet.

Es wurden Vergleichssimulationen durchgeführt, um einerseits zu demonstrieren, dass das angefügte Lotsenmodell an beiden Flughäfen ohne weitere Anpassung arbeiten kann und andererseits das generische Flughafenmodell in der Lage ist, den Verkehr auch auf komplexen Flughäfen problemlos bewegen zu lassen.

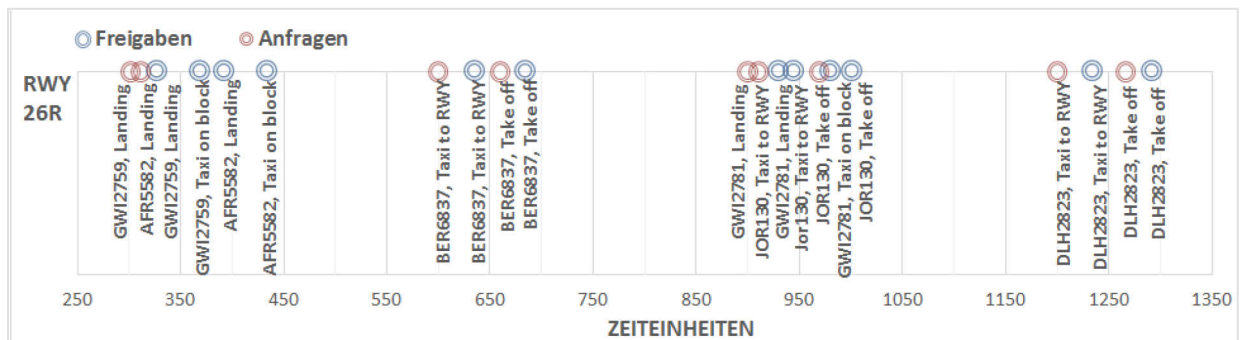


Abbildung 15: Exemplarische Bearbeitung des Verkehrsszenarios durch das Lotsenmodell am Flughafen Braunschweig-Wolfsburg (EDVE)

Die Diagramme in Abbildung 15 und Abbildung 16 veranschaulichen die Bearbeitung des Verkehrs auf dem jeweiligen Flughafen durch das Lotsenmodell. Anfragen sind als rote Punkte dargestellt, Freigaben durch das Lotsenmodell jeweils als blaue. Weiterhin ist zu erkennen, welche Flugzeuge zu welchen Zeiten welche Anfragen stellen, und welche Freigaben das Lotsenmodell erteilt.

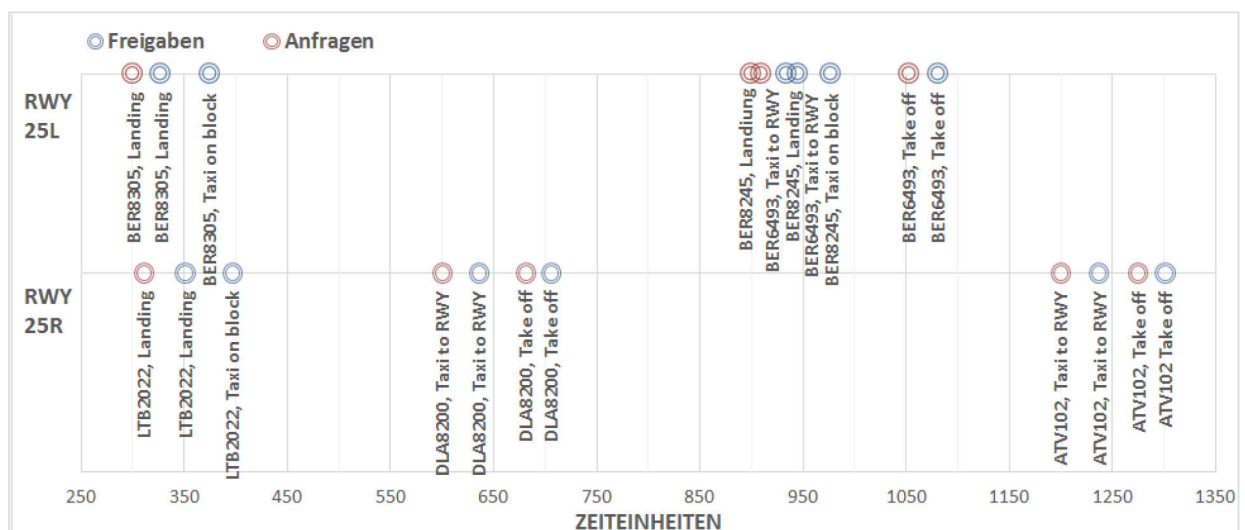


Abbildung 16: Exemplarische Bearbeitung des Verkehrsszenarios durch das Lotsenmodell am Flughafen Berlin-Brandenburg-International (EDDB)

Es ist zu erkennen, dass das Lotsenmodell an beiden Flughäfen eine vergleichbare Verkehrslast bearbeiten kann. In Abbildung 16 ist zu erkennen, dass durch die zur Verfügung stehende zweite Piste die Verteilung des Verkehrs auf beide Bahnen möglich ist, und das Lotsenmodell auch in der Lage ist, eine Verteilung auf beide Bahnen vorzunehmen.

Diskussion und Ausblick

Es wurde gezeigt, dass unterschiedliche Flughäfen durch das generische Flughafen Petri-Netz Modell mit wenig Aufwand umgesetzt werden können. Weiterhin ist das mit dem Flughafenmodell interagierende Lotsenmodell in der Lage, an beiden Flughäfen den Verkehr zu bearbeiten, ohne dass im gesamten Mensch-Maschine-Modell strukturelle Anpassungen nötig sind. Ein Nachteil dieses Ansatzes besteht darin, dass die Rechengeschwindigkeit in der Simulation durch die große Anzahl an Marken im Vergleich zur bisherigen Modellierung herabgesetzt wird. In Zukunft ist es denkbar, auch noch komplexere Flughäfen darzustellen, beispielsweise mit sich kreuzenden Pisten wie in Hamburg (EDDH). Der Ansatz des generischen Flughafenmodells erscheint somit zielführend, um die Weiterentwicklung potenter Lotsenmodelle durch komplexere Anforderungen zu unterstützen. Damit könnten Schwierigkeiten bei der Arbeit in komplexeren Systemen besser identifiziert werden und wertvolle Erkenntnisse zum Verständnis der tatsächlichen Arbeit von Towerlotsen gewonnen werden.

Literatur

- Baumgarten, B. (1996). *Petri-Netze: Grundlagen und Anwendungen* (2nd ed.). Heidelberg ; Berlin ; Oxford: Spektrum, Akad. Verl.
- Cacciabue, P. C. (1998). *Modelling and Simulation of Human Behaviour in System Control*. Springer, London.
- Davidrajuh, R., & Lin, B. (2011). Exploring airport traffic capability using Petri net based model. *Expert Systems with Applications*, 38, 10923–10931.
- Jensen, K. & Kristensen, L. M. (2009). *Coloured Petri Nets: Modelling and Validation of Concurrent Systems*. Berlin, Heidelberg: Springer.
- Moehlenbrink, C., Manske, P. & Kirlik, A. (2012). An analysis of a ground traffic control decision support system based on the 3-step principle of heuristic decision making. In D. de Waard, K. Brookhuis, F. Dehais, C. Weikert, S. Röttger, D. Manzey, S. Biede, F. Reuzeau, & Patrice Terrier (Eds.), *Proceedings of the Human Factors and Ergonomics Society Europe Chapter Annual Meeting*. Toulouse. In press. [Available as open source download]: <http://conference.hfes-europe.org/> [02.08.2013].
- Reisig, W. (2010). *Petrinetze: Modellierungstechnik, Analysemethoden, Fallstudien*: Vieweg + Teubner Verlag.
- Smieszek, H., Manske, P., Hasselberg, A., Rußwinkel, N. & Möhlenbrink, C. (2013). Cognitive Simulation of limited Working Memory Capacity applied to an Air Traffic Control Task. In: R. West & T. Stewart (eds.), *Proceedings of the 12th International Conference on Cognitive Modeling*, Ottawa: Carleton University.

- Werther, B. (2006) Colored Petri net based modeling of airport control processes. In the proceedings of the International conference on Computational Intelligence for Modelling, Control and Automation. CIMCA, Sydney, Australia
- Werther, B., Möhlenbrink, C. & Rudolph, M. (2007). Colored Petri Net Based Formal Airport Control Model for Simulation and Analysis of Airport Control Processes. In *Human Computer Interaction 2007. Human Interface and the Management of Information* (pp. 1027–1036). , 22.-27.7.2007, Beijing, China.

Prospective decision making in a macro-cognitive model of airport traffic control system (MATriCS) based on coloured petri nets

Hardy Smieszek¹ und Fabian Joeres¹

Keywords: macro-cognitive modelling, coloured petri nets, air traffic control, prospective decision making, mental simulation

Abstract

Because of intensively growing volumes of air traffic especially at airports two bottlenecks arise. Besides the capacity of the airport the cognitive capacity of the air traffic controller (ATCO) is a crucial factor. It is therefore of great importance, to understand, how a given traffic scenario at the airport influences the complexity of the task and therefore the cognitive load of the ATCO. Based on basic guidelines for the development of cognitive simulation, a macro-cognitive model was developed enabling investigation of the joint system performance taking cognitive components of the controller into account. Here working memory capacity serves as a measure of cognitive load of the ATCO. In the first version of the model decisions were only based upon information about the current situation. A modelling approach for a new version is presented, taking into account the ability of ATCOs to mentally simulate future states of a traffic situation for making their decisions.

Introduction

At airports air traffic controllers (ATCOs) are responsible for safe and efficient handling of aircraft (AC) within their designated control zone. As the task load at airports varies over the day and the overall amount of air traffic increases, it is of great importance to identify periods of high or low workload, as they are associated with higher risk of human error. By knowing when the controller is at risk to get over-loaded, adaptive computer support could be an opportunity to relieve cognitive load of the controller. It is therefore necessary to investigate the development of workload over time, to identify when the controller is under- or over-loaded.

For the applied problem, a macro cognitive modelling approach was chosen using coloured petri nets (Jensen & Kristensen, 2009) incorporating the environment (airport-process), the interactions (incoming and outgoing information), and the controller's cognitive processes into one model. The cognitive processes were modelled on a low grain level following a macro cognitive approach (Cacciabue & Hollnagel, 1993, see also Smieszek & Rußwinkel, this issue).

¹ Technische Universität Berlin, Marchstrasse 23, 10587 Berlin

By implementing a first simple version of the model (MATriCS v. 1.0; Model of Airport Traffic Control System) a first approximation of workload data is possible and a progression of workload over time for different traffic scenarios at different airports can be generated (Smieszek, Manske, Hasselberg, Russwinkel & Moehlenbrink, 2013).

Nevertheless some processes have to be modelled in more detail to get more reliable and human-like data as well as a more adequate qualitative description of the performance of tower controllers. Especially the decision making process needs to be edited. Here the projection of the development of a particular traffic scenario into the future, as a fundamental ability of air traffic controllers, has to be considered. Until version 1.5 only the actual traffic situation was considered for decision making.

In the next section a brief introduction to the air traffic control task is given. Afterwards the structure of MATriCS in general as well as the decision making process in detail is explained. Subsequently the update version of the decision making process is introduced. Finally the enhancements are discussed.

Air Traffic Control Task

An overview of task analysis studies is provided by Tavanti (2006) which was, among other task analyses, also used for developing a theoretical model of the main tasks of tower controllers (Smieszek, Huber & Juergensohn, 2011).

The main responsibility of tower controllers is to safely and efficiently handle air traffic within their designated control zone. Therefore they have to monitor the progression of traffic on the airport to be able to give clearances to ACs and vehicles to allow them entering specific locations. An arriving AC needs a landing clearance and a taxi clearance to get to its gate. A departing aircraft needs push-back clearance, taxi clearance, and take-off clearance (Werther, 2006).

As a tool for planning, ATCOs use flight strips containing all information about a specific AC. They are printed a few minutes before an AC enters the control zone and have to be updated regularly by the controller as the AC proceeds. All flight strips are deposited and ordered at the strip bay (Smieszek et al., 2013).

MATriCS v.1.5

Version 1.5 of the MATriCS Model is divided into three parts: the airport model, the controller model and the interaction model which provides exchange of information between the airport and the controller model. The airport model was developed in cooperation with the German Aerospace Center (DLR) Braunschweig (see Manske, Smieszek, Hasselberg & Moehlenbrink, this issue). It is designed to support the modelling of various airfields and their different layouts and procedures without much effort. All information needed for interactions of controller and airport are provided through the interaction model. The Controller model is designed

on the basis of action regulation theory following Hacker (2006), Doerner & Schaub (1995) and Frese & Zapf (1994). Four sub-modules are realized representing the individual phases of action regulation: perceive, plan, decide, act. Beyond the sub-modules modelling the action phases a working memory structure is implemented describing a simple single limited capacity system following Miller's 1956 claim of a working memory capacity of "seven plus or minus two chunks". Furthermore the controller is assumed to be familiar with the layout and structure of the specific airport. Hence, information about the positions and lengths of the airport's runways and taxiways, as well as connections between them are easily and correctly retrievable from long term memory. A more detailed description of the action phases and the working memory structure can be found in Smieszek et al. (2013). Regarding the purpose of this paper, a basic understanding of the controller model's decision making process is essential. Therefore the according sub-module is further explained in the following section.

The Decision Making Process v.1.5

Within the decision making process, the controller model evaluates whether a clearance can be given or not and which one should be given. A hypothetical traffic scenario is depicted in figure 1. Here AC 1 is currently taxiing on taxiway segment C2 to its stand. Meanwhile AC 2 requests for a taxi clearance to the runway. The decision process is triggered by the request of AC 2. The controller checks whether the next location intended to be used by AC 2 is currently free or not (taxiway segment G1, figure 1) by taking a look outside onto the airport. Additionally flight strip information about the requesting AC must be obtained from the strip bay. This information is integrated into a decision which leads to the transmission of the entire clearance for AC 2. This already includes all following taxiway segments until the runway, which can easily lead to conflicts (as shown in figure 1, a conflict between AC 1 and AC 2 arises at the intersection of taxiway C3 and G2).

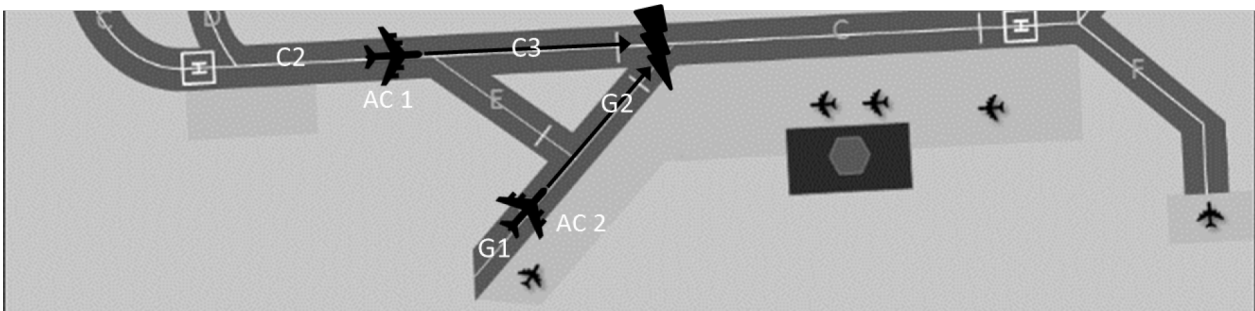


Figure 1: Hypothetical traffic situation indicating the inadequacy of the decision making process of MATriCS v.1.5 for taxiing decisions.

This kind of decision making process is rather simplified as it is solely based upon the current traffic situation and the current status of a specific location. Nevertheless, this representation is sufficient for a first approximation of data and also

sufficient especially for landing and take-off processes, where only information about the status of the runway or runways is required. The runways' states are constantly represented in the controller's mental model and updated when required by the controller's actions or external events.

Decisions concerning taxi clearances, however, have to take into account the state of all taxiway segments that are included in the requested taxiing route. Moreover the controller's judgment is not only affected by the current state of those segments but also by their possible states in the near future. To meet this requirement the controller has to project the current traffic situation into the close future, which covers a duration of the next one to four minutes. This process can be referred to as mental simulation (Klein & Crandall 1995, Klein, Moon & Hoffman, 2006).

Model Extensions

An integration of several task analyses (Smieszek, Huber & Juergensohn, 2011) describes the steps to be taken in these decision and mental simulation processes. To judge the traffic situation the controller has to:

- retrieve the current position and destination of the requesting AC,
- remember and choose a possible route,
- compare the chosen route to taxiing routes of other ACs based on their flight strips and identify ACs whose routes interfere with the chosen route,
- retrieve the positions of those "critical" ACs,
- project the positions of all regarded ACs and decide if any conflicts will occur if a clearance is given for the chosen route,
- decide if a clearance can be given for the entire route, for a part of the route or if at this point no taxi clearance can be given.

The first four steps of mental model updating and refining (Klein, Moon & Hoffman, 2006) can be implemented using the existing cognitive structures of model version 1.5. The mental simulation, summarized in steps 5 and 6, however, includes cognitive processes that have not been regarded yet and will therefore be discussed in further detail.

Cognitive Process and Resources

The controller's decision requires a comparison of the point in time when two relevant ACs reach a certain location. To obtain this information the controller has to estimate the taxiing durations for each AC.

Russwinkel et al. (2011) propose a cognitive mechanism for time estimation that is based on the accumulation of short sequences with a duration of only a few seconds each. The length of those segments is estimated by means of an internal pacemaker. The single segments are split by discrete events and memorized. Long-

er periods of time are then estimated by retrieving and accumulating these segments in the working memory.

Furthermore, Peterken, Brown, & Bowman (1991) hypothesize that a common strategy for the prediction of a disappearing, moving target's position is to keep tracking the target's trajectory visually. Based on these findings it is assumed that the controller integrates both these strategies.

When required to project a current situation into the future, the controller model retrieves the current position and velocity of the previously identified, relevant ACs. The velocity is modelled to be constant and identical for all ACs. To retrieve these positions the controller model needs to look outside the window or at his radar screen, respectively. An average duration of time spans in which the attentional focus lies on the outside view is taken from ATCO observations conducted by Pinska (2008) where a mean scan time of the outside view was measured with 7.2 seconds.

In addition to this information it is assumed that the controller has accurate knowledge of motion durations in the various areas of the airfield based on his experience. Using this knowledge the controller model starts mentally tracking the ACs' motion trajectories. The duration of each motion process is assumed to be represented as proposed by Russwinkel et al. (2011). Therefore the mental simulation of these durations can be modelled to occur analogue to the quoted working memory updating process. However, instead of visually based updating processes the single updating steps are performed by retrieving information from long term memory.

Implementing this strategy in the existing controller model requires estimations and assumptions about the process durations and the required cognitive resources. The process duration can be calculated from the original working memory updating frequency f and the time difference between the current system state and a state in which the controller model can determine if a conflict between two ACs occurs ($t_{conflict}$).

If such a conflict occurs (i.e. two ACs are in one location at the same time) this $t_{conflict}$ can be defined as the time elapsing until the second AC enters the location. If two ACs use the same location at different times (e.g. no conflict occurs but a projection is required to determine this) $t_{conflict}$ is defined as the time elapsing until the first AC leaves the relevant location. The updating frequency results from the lengths of the single time segments as proposed by Russwinkel et al. (2011).

As mentioned before, these segments are split by discrete events whereas the ACs' motion is a continuous process without clearly defined *temporal landmarks*. Hence, the updating frequency f requires further research concerning the time segmentation regarding continuous processes.

However, based on the single $t_{conflict}$ and a frequency f that has yet to be determined the overall number of updates in a system with one requesting AC and m possible conflict ACs can be calculated as:

$$n_{update} = f \cdot \left(\sum_{i=1}^m t_{conflict,i} + \max(t_{conflict,i}) \right) \quad (I)$$

Based on the number of cognitive processing steps n_{update} and the mean visual scanning time of 7.2s the overall time can be determined as:

$$T_{projecting} = 7.2 \text{ s} + n_{update} \cdot t_{update} \quad (II)$$

As explained earlier the updating processes in MATriCS are assumed to be memory processes rather than visual processes. Therefore estimations for the step duration t_{update} are based on the widely accepted cognitive architecture ACT-R (Anderson et al., 2004). In ACT-R, basic memory retrieval processes are hypothesized to last at least 50ms. Their actual duration is dependent on the activation the retrieved chunks currently have in the model's declarative memory (Bothell, 2007). Altering this activation can be used to model different levels of experience with the specific airport layout.

Equation (I) is consistent with the results of Peterken et al. (1991) who found the predicted, invisible length of a moving target's trajectory to have the most severe effect on prediction difficulty. Concerning the cognitive resources occupied and used in this process, particularly visual attention and working memory resources must be regarded. Since the controller model needs to compare the flight strips and retrieve the current motion information of all relevant ACs, the controller's visual attention is modelled to be occupied during the entire process.

In the first step of the decision process the controller model retrieves flight strip information about possible conflict aircrafts. Subsequently the current locations of those vehicles are retrieved from outside view. These information need to be stored and frequently updated in the working memory. When the decision process finishes, the conflict ACs are removed from working memory. The requesting AC, however, is kept in working memory, enabling the controller model to communicate with the AC and check its arrival at the target location after the correspondent taxi duration.

Discussion

The previously proposed model extensions represent an increase of model accuracy regarding the mentioned task analyses and allows for more human-like behaviour of the controller model. It is enabled to judge traffic situations not only based on the current situation but projects the system state into the close future and hereby predicts critical situations. Furthermore the virtual airport safety is consid-

erably increased since ground collisions can now be detected and prevented by the controller model.

However, the implementation of these model extensions requires further fundamental research regarding the working memory updating frequency and general applicability of the WMU hypothesis by Russwinkel et al. (2011) in tasks that concern continuous processes.

Furthermore, the working memory model needs to be updated as well, because the chunking process might show that a single-limited capacity model which is independent from information modality with a capacity of “seven plus or minus two chunks” is insufficient for representing such complex information about AC identities, intentions, positions, and motion patterns.

Literature

- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C. & Qin, Y. (2004). An Integrated Theory of the Mind. *Psychological Review*, 111(4), 1036-1060.
- Bothell, D. (2007). ACT-R 6.0 reference manual. Working draft. From the ACT-R website <http://act-r.psy.cmu.edu/actr6/reference-manual.pdf> [16.07.2013]
- Cacciabue, P. C., & Hollnagel, E. (1995). Simulation of Cognition: Applications. In J.-M. Hoc, P. C. Cacciabue, & E. Hollnagel (eds.), *Expertise. Expertise and technology. Cognition & human-computer cooperation* (pp. 55–73). Hillsdale, N.J: Lawrence Erlbaum Associates.
- Dörner, D. & Schaub, H., 1995. Handeln in Unbestimmtheit und Komplexität. *Organisationsentwicklung*, 14, 34-47.
- Frese, M. & Zapf, D. (1994). Action as the Core of Work Psychology - A German Approach. In H. C. Triandis, M. D. Dunnette, & L. M. Hough, (eds.) *Handbook of Industrial and Organizational Psychology* (pp. 272 – 340). Palo Alto, CA: Consulting Psychologists Press.
- Hacker, W. (2006). Allgemeine Arbeitspsychologie. Psychische Regulation von Arbeitstätigkeiten, Bern: Huber.
- Jensen, K., & Kristensen, L. M. (2009). *Coloured Petri Nets: Modelling and Validation of Concurrent Systems*. Berlin, Heidelberg: Springer.
- Klein, G. A., & Crandall, B. W. (1995). The role of mental simulation in naturalistic decision making. In J. Flach, P. Hancock, J. Caird, & K. Vicente (eds.), *The ecology of human-machine systems* (pp. 324-358). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Klein, G., Moon, B., & Hoffman, R.R. (2006). Making sense of sensemaking 2: A macrocognitive model." *IEEE Intelligent Systems*, 88-92.
- Manske, P., Smieszek, H., Hasselberg, A., & Moehlenbrink, C. (2013). Entwicklung eines generischen Flughafen-Modells für die effizientere Makrokognitive Modellierung des Mensch-Maschine-Systems der Flughafenverkehrs-

- kontrolle mit farbigen Petri-Netzen. In: E. Brandenburg, L. Doria, A. Gross, T. Günzler & H. Smieszek (eds.), *Tagungsband der 10. Berliner Werkstatt Mensch-Maschine-Systeme: Grundlagen und Anwendungen der Mensch-Maschine-Interaktion*, October 10-12, Berlin.
- Peterken, C, Brown, B., & Bowman, K. (1991). Predicting the future position of a moving target. *Perception*, 20, 5-16.
- Pinska, E. (2008). Analysis of Visual Attention in Augmented Reality Environments for Airport Tower Control (Dissertation). École Pratique des Hautes Études.
- Russwinkel, N., Urbas, L., & Thüring, M. (2011). Predicting temporal errors in complex task environments: A computational and experimental approach. *Cognitive Systems Research*, 12(3-4), 336-354.
- Smieszek, H. (2013). Micro-cognition and Macro-cognition: Trying to bridge the gap. In: E. Brandenburg, L. Doria, A. Gross, T. Günzler & H. Smieszek (eds.), *Tagungsband der 10. Berliner Werkstatt Mensch-Maschine-Systeme: Grundlagen und Anwendungen der Mensch-Maschine-Interaktion*, October 10-12, Berlin.
- Smieszek, H., Huber, S., & Juergensohn, T. (2011). Ein Integratives Prozessmodell der Hauptaufgaben von Rollverkehrslotsen als Grundlage für Modellierung und Simulation. In S. Schmid, J. Adenauer, M. Elepfandt, & A. Lichtenstein (eds.), *Tagungsband der 9. Berliner Werkstatt Mensch-Maschine-Systeme: Reflexionen und Visionen der Mensch-Maschine-Interaktion*, October 5-7, Berlin, Düsseldorf: VDI Verlag.
- Smieszek, H., Manske, P., Hasselberg, A., Russwinkel, N., & Moehlenbrink, C. (2013). Cognitive Simulation of limited Working Memory Capacity applied to an Air Traffic Control Task. In: R. West & T. Stewart (eds.), *Proceedings of the 12th International Conference on Cognitive Modeling*, Ottawa: Carleton University. [Available as open source download]: <http://iccm-conference.org/2013-proceedings/170/0170-paper.pdf> [07.08.2013]
- Tavanti, M. (2006) Control Tower Operations: A literature review of task analysis studies. Eurocontrol Experimental Center, EEC Report 2006-6.
- Werther, B. (2006) Colored Petri net based modeling of airport control processes. In *Proceedings of the International conference on Computational Intelligence for Modelling, Control and Automation*. CIMCA, Sydney, Australia.

Supervising the Supervisor: An Approach to Improve Scanning Behaviour during UAV Supervisory Control

Florian Fortmann¹ and Bertram Wortelen

Keywords: Situation Awareness, Supervisory Control Task, Unmanned Aerial Vehicles, Eye Tracking

Abstract

Progressive automation in the aviation domain has increasingly changed the role of human operators from manual control to supervisory control and enabled human operators to handle complex technical systems beyond limited human capabilities. However, this benefit did not come without a price. Inadequate scanning behaviour during the execution of supervisory control tasks can reduce situation awareness (SA) leading to a new type of human error and even accidents. In this paper we present the assistant system Supervisory Guide, which aims at improving the scanning behaviour of a human operator supervising a swarm of highly autonomous unmanned aerial vehicles (UAV) by guiding visual attention to relevant but unattended information sources on the user interface. We present results of an experiment performed with 15 subjects to evaluate the effect of Supervisory Guide on scanning behaviour, SA and workload. Overall, the results show that scanning behaviour and SA significantly improve, while workload increases.

Introduction

SA is defined as a state of knowledge including the perception of the elements in the environment (level 1), the comprehension of their meaning (level 2) and the projection of their status in the near future (level 3) (Endsley, 1995). The qualities of higher levels of SA depend on the qualities of lower levels (level 1 → level 2 → level 3). Thus, the essence of building overall good SA in working environments mainly composed of visual user interfaces is the execution of adequate scanning behaviour as a means to build SA on level 1. The effect of reduced SA on any level is that human errors expressed by taking an incorrect course of actions in a given situation are more likely to occur. Studies performed on the basis of reports from the Aviation Safety Reporting System have shown that more than 70% of human errors in the aviation domain are caused by insufficient SA (Jones & Endsley, 1996). Furthermore, 80% of these human errors can be attributed to level 1 (Jones & Endsley, 1996).

An emerging field within the aviation domain is the operation of unmanned aerial vehicles (UAV). Today, a highly automated UAV is typically operated via dedicated user interfaces (UI) by a team of human operators (many-to-one). However,

¹ OFFIS - Institute for Information Technology, Oldenburg, Germany

the interest of stakeholders in UAV swarms, which can be operated via a dedicated UI by a single human operator (one-to-many) increased significantly (Brian, McLaughlan, & Baker, 2005). The expected benefits of these one-to-many unmanned aerial systems are, e.g., reduced costs for personnel and higher system reliability and resilience. Due to the high levels of UAV automation human operators are in charge of supervisory control tasks (SCT). Basically, UAV supervisory control means that the human operator has to continuously supervise the UAV and environment parameters in order to detect and handle anomalies. The complexity of SCTs in one-to-many systems is very high because the human operator has to oversee a multitude of information in a dense, dynamic information space. Therefore, the execution of adequate scanning behaviour is very important for the safe operation of the UAVs.

This paper presents Supervisory Guide - an assistant system which aims at improving the scanning behaviour of a human operator in charge of supervising a swarm of UAVs by guiding his/her visual attention on the UI using dynamic, visual cues. The invocation of cues depends on the detection of relevant but unattended information elements – called information gaps. The detection of information gaps is based on two inputs: (1) rules are applied to identify relevant information elements in a given situation; (2) scanning behaviour (recorded by an eye tracker) is analysed to estimate the mental SA state of the human operator. The estimation of the mental SA state focuses on two characteristics of scanning behaviour: (1) glances on and (2) diversions from information sources. We present results of an experiment performed with 15 subjects to evaluate the effect of Supervisory Guide on scanning behaviour, SA and workload. Overall, the results show that scanning behaviour and SA significantly improve, while workload increases.

Characteristics of Scanning Behaviour

Eye movement analysis is an important instrument for the analysis of internal human states. With the low priced availability of eye tracking systems, the analysis of scanning behaviour has become a common task when studying user behaviour. However it is not always an easy one. Eye movements can be characterized in many different ways and provide many different kinds of information. Furthermore different methods are available to assess these kinds of information. For the interpretation and replication of the results of an eye tracking study a precise description of the utilized methods and measures are required. If not mentioned otherwise all terms used in the present work, that describe properties of the eye movement behaviour, are used as it is defined in ISO 15007-1 (ISO, 2002).

Typically the first step in eye movement analysis is the pre-processing of raw eye tracking data. This involves the removal of undefined data points, which mainly results from blinks, and the subsequent classification of data segments. If the information sources, that the user utilizes, are not known a priori, the classification typically differentiates between fixations and saccades. Approaches like Velocity

Threshold Identification or Dispersion Threshold Identification are simple ways to identify fixations (Salvucci, 1999). However in most everyday environments the allocation of visual attention is mainly driven by top down factors (Land & Tatler, 2009), and is thus strongly constrained by the task that is executed. In these cases all task-relevant information sources are typically known beforehand. The analysis of eye movements is then used to identify to which of these information sources the user is looking. It is of less interest, how many consecutive fixations are directed to one and the same information source. Therefore eye tracking data segments are classified on the basis of glances. A glance is defined “from the moment at which the direction of gaze moves towards a target [...] to the moment it moves away from it.” (ISO, 2002). Thus a glance includes the transition from the previous information and the dwell time during which the gaze keeps located within the information source. This is reasonable, because in top-down attention processes the mental focus of attention shifts before the visual focus does. Thus the transition to an information source should be grouped to the glance while the transition away from it should not.

The second typical step in eye movement analysis is the calculation of aggregated measures and their interpretation for the investigated situation. Very common measures are percentage dwell times (PDTs), mean glance durations and glance frequencies. But as mentioned in the introduction, the objective of Supervisory Guide is to avoid long periods of time, in which an information source receives no attention (duration of diversion (ISO, 2002)). In order to assess the improvement in scanning behaviour that is achieved by using Supervisory Guide, it is thus more important to identify extreme behaviour than to use measures of mean values

The measure of scanning behaviour that Supervisory Guide shall optimize is defined as follows. Within an observed time interval $[0..T]$ all durations of diversions for each information source are calculated. The i -th duration of diversion for information source A is denoted by $\Delta_{A,i}$, the number of diversion periods for an information source α is denoted by n_α and the number of information sources by N . In order to penalize long durations of diversion, the scanning measure calculates the root mean square over all durations of diversions within the time interval T :

$$\text{RMSDoD} = \sqrt{\frac{\sum_{o \in O} \sum_{i=1}^{n_o} \Delta_{o,i}^2}{\sum_{o \in O} n_o}} \quad (1)$$

Supervisory Guide

Previously performed experiments have shown that eye movements are valuable indications of SA - especially in the aviation domain (Moore, 2009; Ratwani, McCurry, & Trafton, 2010; van de Merwe, van Dijk, & Zon, 2012). Supervisory Guide focuses on two fundamental cognitive processes related to eye movements: (1) memorization of information due to glances on information sources and (2) forgetting of information due to long diversions from information sources. There-

fore, Supervisory Guide analyses eye movements on the UI recorded by an eye tracker in context of the information visually displayed on the UI. The UI used for UAV swarm supervision mainly consists of a 2-D map displaying the state of information objects (e.g., UAVs). Formally, we denote the overall number of information objects as a set: $O = \{o_1, o_2, \dots, o_n\}$. Each information object is associated with a number of information elements (e.g., UAV.location). Formally, we denote the overall number of information elements as a set: $E = \{e_1, e_2, \dots, e_m\}$. The subset of information elements, which the human operator can perceive by scanning an information object with the eyes is defined by the function $f: O \rightarrow \mathcal{P}(E)$. Supervisory Guide maps sequences of raw eye tracking data points lasting longer than 200 ms on an information object o to a valid glance φ_o . Based on the detection of glances, Supervisory Guide continuously updates a mental SA model $M = \{\mu_{e_1}, \mu_{e_2}, \dots, \mu_{e_m}\}$. Each element $\mu_e \in M$ represents the current mental state of an information element $e \in E$ and is defined as a tuple $\mu_e = \langle v_e, p_e \rangle$, where v_e is the perceived value of e and p_e is the current perception level ($p \in [0,1]$) of e . Supervisory Guide continuously updates M by overriding values and increasing perception levels of focused elements, and decreasing perception levels of unfocused elements.

Based on the state of the mental SA model, Supervisory Guide detects information gaps. We define an information gap as a deviation between (1) what the human operator should be aware of and (2) what he/she is actually aware of. The first aspect is determined by pre-defined rules mapping environment states to relevant information elements. The second aspect is determined by the mental SA model. Supervisory Guide considers two types of information gaps. The first type refers to the distance δ_{v_e} between v_e and the actual value of e . The second type refers to the distance δ_{p_e} between p_e and the maximum perception level. In this paper, we applied a threshold satisfaction approach using two thresholds θ_1 and θ_2 to distinct three criticality levels (safe, caution and danger) for δ_{v_e} and δ_{p_e} . During an expert evaluation a mapping of criticality levels to visual cues and thresholds was defined (Frische, 2013). The result of the evaluation was that no visual cue should be invoked for the criticality level "safe", yellow color should be used for the level "caution" and red should be used for the level "danger". The selection of this mapping is also supported by (Salvendy, 2012). The mapping is depicted in Tab.1:

Tab.1: Mapping of criticality levels to thresholds and visual cues.

Criticality Level	v_e		p_e	
	Thresholds	Colour	Thresholds	Colour
Safe	$\delta_{v_e} \leq \theta_{1_{v_e}} < \theta_{2_{v_e}}$	No	$\delta_{p_e} \leq \theta_{1_{p_e}} < \theta_{2_{p_e}}$	No
Caution	$\theta_{1_{v_e}} < \delta_{v_e} \leq \theta_{2_{v_e}}$	Yellow	$\theta_{1_{p_e}} < \delta_{p_e} \leq \theta_{2_{p_e}}$	Yellow
Danger	$\theta_{1_{v_e}} < \theta_{2_{v_e}} < \delta_{v_e}$	Red	$\theta_{1_{p_e}} < \theta_{2_{p_e}} < \delta_{p_e}$	Red

On the UI, information objects are surrounded with rectangular shapes coloured based on the most critical variable (v_e or p_e). An example is shown in Fig.1.



Figure 7: User Interface for supervisory control of a highly automated UAV swarm. Supervisory Guide triggers visual cues based on the detection of information gaps and the classification of criticality levels.

Experiment

The goal of this experiment was to evaluate the impact of Supervisory Guide on visual attention allocation, SA and workload. In total, 15 students (female = 3) from the University of Oldenburg participated for monetary compensation. Their age ranged from 21 to 31 (\bar{x} =25.1, σ =2.46). We asked the students for their level of experience in playing strategy games on the computer because the scenarios and the UI are comparable to popular strategy games. On a scale ranging from 1 (= no experience) to 5 (lots of experience) the average rating on game playing experience was \bar{x} =3.73 (σ =.93). None of the students needed glasses.

Within the scenarios the subjects had to supervise a swarm of three highly automated UAVs to extinguish fires in an area which was not accessible by human fire fighters using ordinary equipment. In order to extinguish the fire sources each UAV had to load water at a water source and unload water at one of the fire locations. During mission execution each UAV had to frequently recharge energy at a base. The UAVs selected tasks autonomously. The subjects could not intervene.

However, at random times two distinct types of events were induced into the scenarios, which the subjects had to handle manually. First, UAV malfunctions were triggered which had to be detected and resolved by the subjects by clicking the UAV icon and then pressing the „Repair UAV“ button on the UI. Second, intruders entered the mission area, which had to be classified (using the aircraft (a/c) list on the UI). Intruders could either be firefighting aircrafts or civil aircrafts. Firefighting aircrafts had to be allocated to the fire sources by clicking the aircraft icon, a target fire and then pressing the “Allocate Target” button. Civil aircrafts had to be detoured by clicking the aircraft icon and pressing the “Detour Flight” button.

A notebook was used to run the simulation, the UI and Supervisory Guide. The eye movements of the subjects were recorded with the Dikablis head mounted eye tracking system from Ergoneers. A 26-inch monitor was used to display the UI. Two eye tracking markers were attached to the monitor. The subjects were seated approximately 0.6 meters from the monitor and used a mouse device to interact with the UAVs. The experiment setup is depicted in Fig.2.

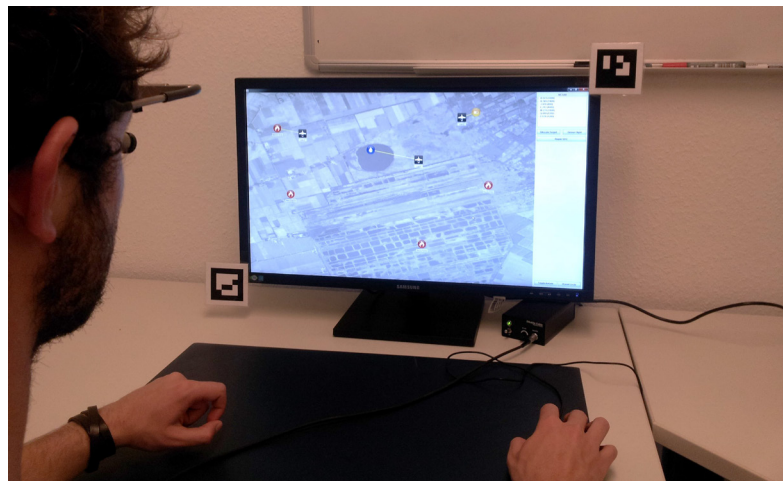


Figure 2: The experiment setup.

A repeated measures design was selected. The experiment consisted of two comparable scenarios (control and experiment condition), where the subjects had to supervise a swarm of highly automated UAVs. The independent variable was the assistance mode: in the control condition Supervisory Guide was deactivated and in the experiment condition Supervisory Guide was activated. The dependent variables were: (1) scanning behaviour, (2) SA and (3) workload. We counterbalanced the order of conditions to avoid order effects, e.g., due to learning and fatigue. The overall duration of the experiment was approx. 75 minutes per subject: 15 minutes introduction + 25 minutes for each condition + 10 minutes for a final interview.

At the beginning of the experiment each subject had to sign a letter of consent. After the letter was signed, each subject was introduced to UAVs and the UAV domain. Then, we presented the UI, explained the objective of the mission and the functions of the UI. If the first scenario was the experiment condition we also ex-

plained the function of Supervisory Guide, otherwise we explained the assistant system prior to the second experiment. Then, a short familiarization scenario was run. After each subject was able to perform the mission, each subject was calibrated for the eye tracker and started the scenarios. Immediately after each scenario the monitor was blinded and two questionnaires (see section "Measures") were handed out to the subjects. After both scenarios were completed a semi-structured interview was performed.

The data of all 15 subjects was analysed with regard to scanning behavior, SA and workload. The RMSDoD was calculated as described in this paper to evaluate whether the Supervisory Guide improved the scanning behaviour. The experimental data showed a significant improvement in scanning behaviour, when using the Supervisory Guide. For each participant the RMSDoD was calculated for the assisted and unassisted condition. The RMSDoD was significantly ($p < .01$) smaller in the experiment condition ($\bar{x} = 25.05s$, $\sigma = 1.84s$) then in the control condition ($\bar{x} = 27.94s$, $\sigma = 2.24s$). The results are shown in Fig.3 (left).

The NASA Task Load Index (NASA-TLX) questionnaire was used to evaluate workload of the subjects. NASA-TLX is a widely accepted technique for workload evaluations cited in hundreds of papers (Hart, 2006). We applied the raw version of the NASA-TLX, which does not perform an individual weighting of the workload dimensions. According to (ref), this is a legitimate approach and does not affect the results. The analysis of the NASA-TLX questionnaires reveals a score of $\bar{x} = 5.52$ ($\sigma = 2.21$) for the control condition and a score of $\bar{x} = 6.86$ ($\sigma = 2.5$) for the experiment condition (see Fig.3 (right)). The differences between the control and experiment conditions were significant ($p < .05$).

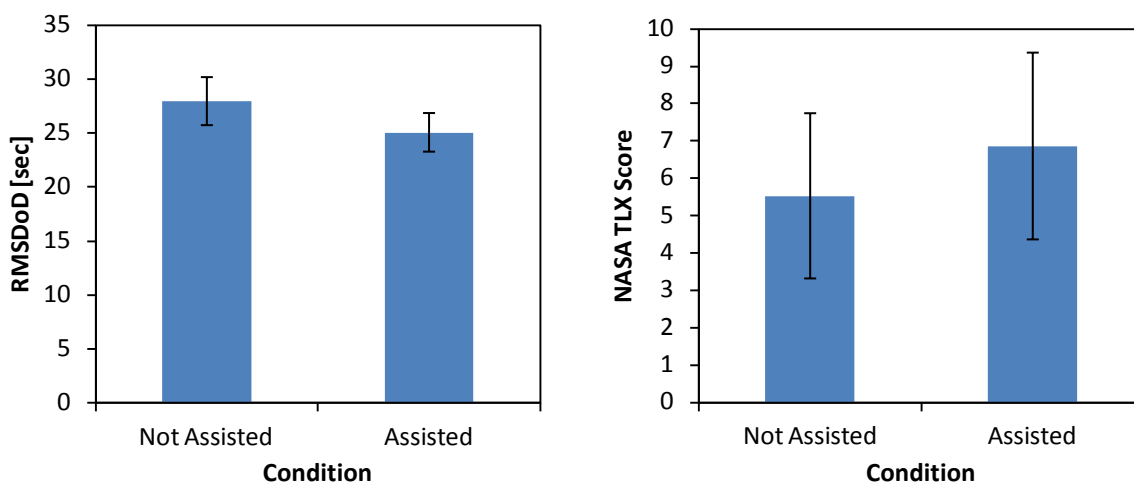


Figure 3: Results of the RMSDoD (left) and NASA-TLX (right).

The Situation Awareness Rating Technique (SART) was used to evaluate the overall SA of the subjects. SART is the most widely known SA self-rating technique. We used a shortened version of the technique (SART-3), which consisted of three statements. Each of the statements had to be answered on a Likert scale ranging from 1 (strongly agree) to 5 (strongly disagree). The statements were: “The

task was mentally demanding”, “I could have managed more UAVs” and “I was able to keep track of what was going on”. The analysis of the SART-3 questionnaires was performed separately for each question. The average score for question 1 was $\bar{x}=3.27$ ($\sigma=.68$) for the control condition (not assisted) and $\bar{x}=2.27$ ($\sigma=1$) for the experiment condition (assisted). The average score for question 2 was $\bar{x}=1.53$ ($\sigma=.5$) for the control condition and $\bar{x}=2.27$ ($\sigma=.77$) for the experiment condition. The average score for question 3 was $\bar{x}=1.93$ ($\sigma=.68$) for the control condition and $\bar{x}=1.4$ ($\sigma=.49$) for the experiment condition. The results are also shown in Fig.4. For all three questions, the differences between the control and experiment conditions were significant ($p<.01$).

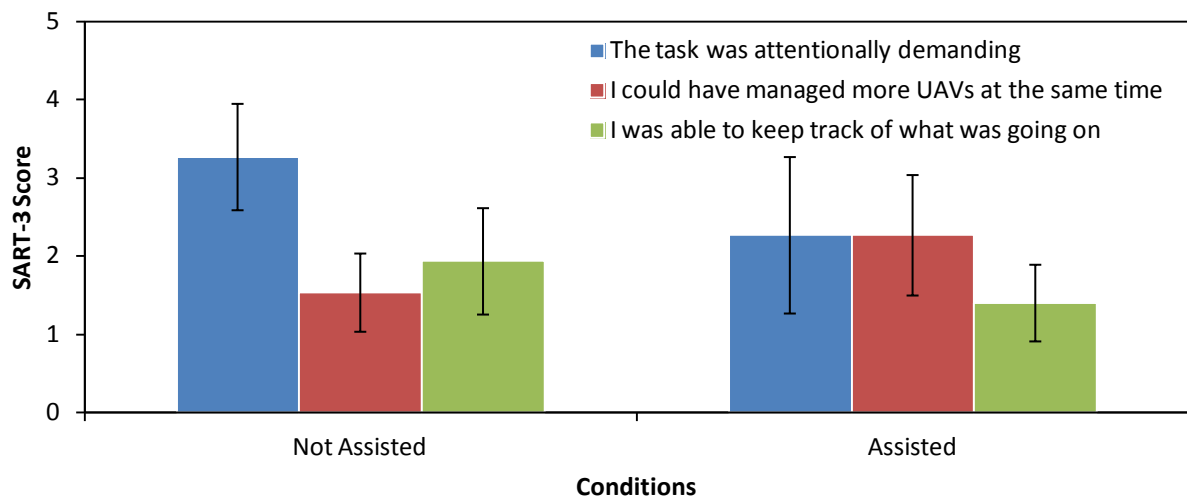


Figure 4: Results of the SA evaluation with SART-3.

Finally, a semi-structured interview was performed to gather further valuable feedback on Supervisory Guide. Overall, the subjects mentioned that they had problems with wearing the eye tracking device for long periods of time. Reasons were the weight of the device and the camera, which irritated the eye field. Nonetheless, the subjects agreed that they felt safer with the assistant although they also felt a more stressed due to the obtrusiveness of visual cues. Most of the subjects believed that Supervisory Guide could actually improve SA as it worked like a virtual co-operator.

Conclusions

This paper presented results of an experiment performed with Supervisory Guide – an assistant system for improving the scanning behavior of human operators in charge of UAV swarm supervisory control. Supervisory Guide implements an SA-based strategy for guiding visual attention, which is based on the analysis of glances and diversions. If Supervisory Guide detects information gaps, then visual cues are invoked on the UI resulting in an immediate shift of visual attention to relevant but unattended information.

Our results show that Supervisory Guide significantly reduces the duration of diversions resulting in a much more fluent and regular shift of glances between the different information sources. Further, the results show that SA improved significantly, when Supervisory Guide was activated although workload increased. In the future, we will investigate whether it is possible to apply more unobtrusive visual cues, e.g., by modifying the transparency and size of cues. Then, it has to be evaluated which effect unobtrusive cues have on SA. Further, we are also interested in comparing the SA-based invocation of visual cues against a more random strategy in order to assess the quality of the information gap detection and its value within the assistant.

Literature

- Brian, H. H., McLaughlan, B., & Baker, M. (2005). Swarm Control in Unmanned Aerial Vehicles. In *Proceedings of International Conference on Artificial Intelligence (IC-AI)*, CSREA. Press.
- Endsley, M. R. (1995). Toward a Theory of Situation Awareness. *Human Factors*, 37(1), 32–64.
- Frische, F. (2013). SA-Based Guidance to Aid UAV Swarm Supervisory Control: What do Experts Say? In *Proceedings of Mensch und Computer 2013*.
- Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, pp. 904–908).
- ISO. (2002). ISO 15007-1:2002. Part 1: Road Vehicles: Measurement of Driver Visual Behaviour with Respect to Transport Information and Control Systems -- Definitions and Parameters. Genf, Switzerland: ISO International Organization for Standardization.
- Jones, D. G., & Endsley, M. R. (1996). Sources of situation awareness errors in aviation. *Aviation space and environmental medicine*, 67(6), 507–512.
- Land, M., & Tatler, B. (2009). *Looking and Acting: Vision and eye movements in natural behaviour* (p. 320). Oxford University Press, USA.
- Moore, K. S. (2009). Comparison of Eye Movement Data to Direct Measures of Situation Awareness for Development of a Novel Measurement Technique in Dynamic, Uncontrolled Test Environments. Clemson University.
- Ratwani, R. M., McCurry, J. M., & Trafton, J. G. (2010). Single operator, multiple robots: An eye movement based theoretic model of operator situation awareness. *2010 5th ACM/IEEE International Conference on Human Robot Interaction HRI*, (55), 235–242. doi:10.1109/HRI.2010.5453191
- Salvendy, G. (2012). *Handbook of human factors and ergonomics*. Wiley.
- Salvucci, D. D. (1999). *Mapping eye movements to cognitive processes*. Carnegie Mellon University.

Van de Merwe, K., van Dijk, H., & Zon, R. (2012). Eye Movements as an Indicator of Situation Awareness in a Flight Simulator Experiment. *The International Journal of Aviation Psychology*, 22(1), 78–95.

Mensch-Maschine-Interaktion im Betrieb der SBB

Jonas Brüngger¹, Peter Grossenbacher², Pia Zwahlen², Jasmin Zimmermann¹, Kathrin Gärtner¹, Toni Wäfler¹ und Katrin Fischer¹

Schlüsselwörter: Automatisierung, Zugverkehrsleitung, Funktionsallokation

Zusammenfassung

Im Projekt „Mensch-Maschine-Interaktion im Betrieb der SBB“ geht es um die Fragen, welche Auswirkungen Veränderungen der Automatisierung auf die Tätigkeiten in der Zugverkehrsleitung haben und welche Anforderungen daraus für die Gestaltung der betroffenen Systeme entstehen. Erste Ergebnisse zum aktuellen Stand der Automatisierung verschiedener Arbeitstätigkeiten in der Zugverkehrsleitung liegen vor.

Im „Werkstatt Track“ soll den Fragen nachgegangen werden, mit welchen Methoden die Auswirkungen von Automatisierung auf ein soziotechnisches System am besten antizipiert und beurteilt werden können, welche arbeitspsychologischen Kriterien für die Bewertung relevant sind und wie diese Kriterien geeignet operationalisiert werden können. Darüber hinaus soll diskutiert werden, ob Tätigkeiten mit spezifischen Automatisierungsprofilen (in Anlehnung an Parasuraman, Sheridan & Wickens, 2000) auch mit spezifischen Risiken verbunden sind.

Ausgangslage

Durch technische und organisationale Veränderungen bei den Schweizerischen Bundesbahnen (SBB) wird sich in den nächsten Jahren der Grad der Automatisierung bei unterschiedlichen Arbeitstätigkeiten in der Zugverkehrsplanung, -leitung, Disposition und Fahrgastinformation (nachfolgend unter Zugverkehrsleitung zusammengefasst) weiter erhöhen. Dadurch stellen sich zentrale Fragen, die für die Sicherheit, Zuverlässigkeit und Wirtschaftlichkeit der Zugverkehrsleitung in Zukunft von entscheidender Bedeutung sein werden. In einem Forschungsprojekt mit der Fachhochschule Nordwestschweiz, den SBB und IBM Rüschlikon sollen die folgenden Fragen beantwortet werden:

- Welche Auswirkungen – positive und negative – hat die Automatisierung unterschiedlicher Tätigkeiten auf die Arbeit der Zugverkehrsleitung, sowohl im Regelbetrieb als auch im Störfall?
- Welche Anforderungen resultieren daraus für die künftige Gestaltung der Technik, der Arbeitsorganisation, der Aufgabenverteilung sowie für Schulung und Training, um negative Auswirkungen zunehmender Automatisie-

¹ Fachhochschule Nordwestschweiz FHNW, Hochschule für Angewandte Psychologie, Olten

² Schweizerische Bundesbahnen SBB

rung aufzufangen und die Sicherheit und Zuverlässigkeit des Systems zu gewährleisten?

Im Allgemeinen soll die Automatisierung Fehler vermeiden, oft schafft sie jedoch auch neue. Künftig wird deshalb in der Zugverkehrsleitung eine neue Qualität der Automatisierung erforderlich sein, die nicht primär von den technischen Möglichkeiten, sondern verstärkt von den Fähigkeiten und Bedürfnissen des Menschen ausgeht, um mögliche negativen Folgen der Automatisierung zu vermeiden. Zur Beantwortung der Fragestellungen wurde das Projekt in zwei Phasen gegliedert: in der ersten Phase wurden Daten zu den Tätigkeiten und der aktuellen Gestaltung der Automatisierung in der Zugverkehrsleitung erhoben. Neben der Beschreibung von Systemeigenschaften und kognitiven Anforderungen an die Zugverkehrsleitung wurden dazu tätigkeitsspezifische Profile des Automatisierungsgrades erstellt. In einer nächsten Phase des Projekts soll eine Abschätzung der künftigen technischen Entwicklungen bei den SBB und deren Einfluss auf das soziotechnische System der Zugverkehrsleitung stattfinden.

Theoretischer Hintergrund

In Anlehnung an Hauss und Timpe (2000) wird von Automatisierung oder Automation gesprochen, wenn einzelne Funktionen oder ganze Tätigkeiten, die bisher von Menschen ausgeführt wurden, auf Maschinen übertragen werden. Unter Automation wird das Ergebnis einer solchen Übertragung verstanden.

In einem Modell der Automation von Parasuraman, Sheridan & Wickens (2000) werden verschiedene Arten und Stufen der Automation nach zwei Aspekten unterschieden (Model for Types and Levels of Automation). Zum einen wird nach der Art der automatisierten Funktion differenziert: (1) Informationsaufnahme, (2) Informationsverarbeitung, (3) Entscheidungsfindung und (4) Handlungsausführung. Zum anderen wird nach dem Ausmass der jeweiligen Automation unterschieden. Dabei wird untersucht, in welchem Umfang Funktionen an Maschinen übertragen wurden. Im Sinne dieses Modells können aufgabenspezifische Automatisierungsprofile abgebildet werden (siehe Abb.1). Parasuraman et al. (2000) schlagen vor, beim Design von Systemen darauf zu achten, welche Wirkungen verschiedene Automatisierungsgrade der vier vorgeschlagenen Funktionen auf menschliche Leistungskriterien, die Zuverlässigkeit der Automation sowie die Risiken (Costs of Decision/Action Outcomes) haben.

Ein durchdachtes Design der Automatisierung von Tätigkeiten soll dazu dienen, typische Fehler, die bei zunehmender Automatisierung auftreten können, zu vermeiden. Zu denken sei hier u. a. an die sogenannten „Ironien der Automatisierung“ (Bainbridge, 1983). Manzey (2012) fasst drei hauptsächliche unerwünschte Effekte der Automatisierung zusammen. Diese werden im Folgenden näher erläutert.

Verlust des Situationsbewusstseins

Der Begriff „Situationsbewusstsein“ (engl. Situation Awareness, Endsley, 2000) bezieht sich auf die Wahrnehmung und das Wissen eines Operateurs (z.B. eines Piloten oder eines Zugverkehrsleitenden) über den Zustand des zu kontrollierenden Systems und seiner Objekte.

Endsley (2000) unterteilt Situation Awareness in drei Ebenen:

- Ebene 1: Wahrnehmung des aktuellen Zustandes der Systemparameter.
- Ebene 2: Verständnis der Bedeutung der Systemparameter.
- Ebene 3: Vorhersage (Antizipation) des zukünftigen Zustandes des Systems.

Situationsbewusstsein setzt aufgabenrelevantes Wissen und die Fähigkeit voraus, aus diesem Wissen künftige Systemzustände vorherzusagen. Für Zugverkehrsleitende, welche das dynamische System „Züge - Gleisanlage - Fahrplan - Menschen - Situation“ handhaben müssen, sind alle drei Ebenen relevant. Ein Verlust des Situationsbewusstseins droht bei unangemessener Automatisierung dann, wenn Rückmeldekanäle verändert oder eingeschränkt werden, aber auch bei Intransparenz der Systemfunktionen, bei zu hoher Systemkomplexität, bei mangelndem Systemverständnis und bei übersteigertem Vertrauen in die Automation (Manzey, 2012).

Verlust von Fertigkeiten

Zunehmende Automatisierung kann bei den Operateuren zu einem Verlust an Fertigkeiten führen, wenn diese nie oder nur noch selten die Gelegenheit haben, ein System manuell zu bedienen. Bereits antrainierte Fähigkeiten können über die Zeit schlechter werden oder sogar ganz verloren gehen (deskilling, Manzey, 2012). Dies ist für die Aufgabenausführung insbesondere dann problematisch, wenn ein Operateur bei Automationsfehlern oder -ausfällen in der Lage sein muss, automatisierte Funktionen wieder zu übernehmen und selbst auszuführen (Parasuraman, et al., 2000).

Übersteigertes oder zu geringes Vertrauen in die Automation

Eine hohe wahrgenommene Zuverlässigkeit der Automation, hohe Beanspruchung der Operateure durch gleichzeitige Aufgaben (Multi-Tasking), Müdigkeit, geringes Vertrauen in die eigene Leistungsfähigkeit sowie mangelnde eigene Erfahrung mit Automationsfehlern können dazu führen, dass Operateure ein übersteigertes Vertrauen in die Automation entwickeln (Manzey, 2012) und sich zu sehr auf die Technik verlassen (complacency, Parasuraman, Molloy & Singh, 1993). In der Folge werden automatisierte Funktionen mangelhaft oder nachlässig überwacht. Dies kann zu Unterlassungs- und Handlungsfehlern führen.

Mangelndes Vertrauen in die Automation kann ebenfalls dysfunktional sein, wenn es zur Unterschätzung der „wahren“ Systemzuverlässigkeit führt und damit

automatisierte Systeme nur unzureichend genutzt werden. Dies ist besonders problematisch bei Alarm- und Warnsystemen (Parasuraman & Riley, 1997; Meyer, 2001).

Methodisches Vorgehen

Das Projekt befindet sich derzeit in seiner ersten Phase, in der die Aufgaben und kognitiven Anforderungen der Zugverkehrsleitung in Bezug zum derzeitigen Stand der Automatisierung beschrieben werden. Hierzu wurden Dokumentenanalysen und leitfadengestützte Tätigkeitsbeobachtungen sowie Interviews durchgeführt. Die Leitfäden wurden auf Basis der KOMPASS Methode zur Bewertung und Gestaltung von Produktionssystemen (Wäfler, Windischer, Ryser, Weik & Grote, 1999) erstellt. Insbesondere lagen dabei die (1) Qualität der zur Verfügung stehenden Daten und Informationen, (2) Prozesstransparenz, (3) Kopplung und (4) der Automatisierungsgrad in den auszuführenden Aufgaben sowie (5) Störungen und Schwankungen im Fokus. Ergänzt wurden die Leitfäden im Hinblick auf die Methoden der Cognitive Task Analysis (Militello & Hutton, 1998). Neben Aufgabenbeschreibungen mit Fokus auf kognitive Anforderungen wurden daraus Profile des Automatisierungsgrades der verschiedenen Tätigkeiten in der Zugverkehrsleitung erstellt (siehe Abb.1).

Diese Profile bilden jeweils tätigkeitsspezifisch ab, in welchem Masse die Informationsaufnahme, -verarbeitung, Entscheidungsfindung und Handlungsausführung derzeit automatisiert sind. Die Kriterien zur Beurteilung der Automatisierungsgrade wurden aus den Beschreibungen zum „Modell verschiedener Arten und Stufen der Automation“ von Parasuraman et al. (2000) entnommen.

Ergebnisse

Aus den Automatisierungsprofilen wird deutlich, dass sich sowohl die untersuchten Funktionen (Zugverkehrsplaner, Zugverkehrsleitende, Disponenten, Disponenten Assistenten und Informationsspezialisten) wie auch die unterschiedlichen Tätigkeiten dieser Funktionen zum Teil stark voneinander unterscheiden. Die verschiedenen Tätigkeiten bieten dementsprechend auch individuelle Herausforderungen im Umgang mit automatisierten Systemen.

Zur Veranschaulichung wird beispielhaft das Profil der Zugverkehrsleitenden mit den Tätigkeiten Überwachung/Disposition und Störungsbehandlung (Abb. 1) beschrieben.

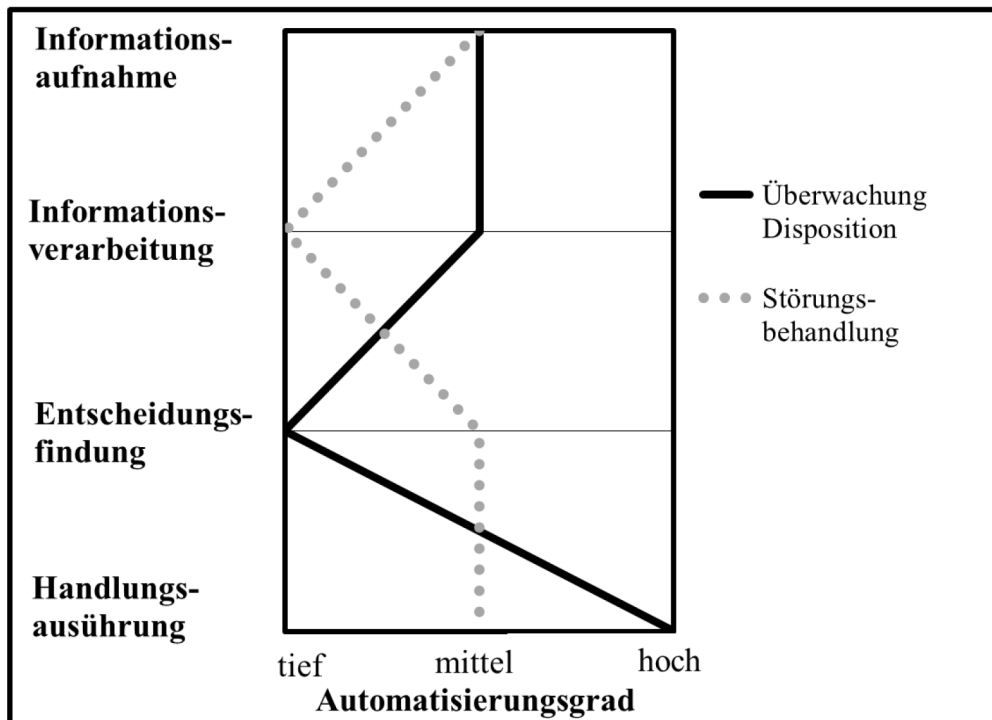


Abb.1: Bsp. Automatisierungsprofil eines Zugverkehrsleitenden für die Tätigkeiten Überwachung/Disposition und Störungsbehandlung (in Anlehnung an Parasuraman et al., 2000)

Die Zugverkehrsleitenden nehmen Informationen sowohl über den Computer als auch über verschiedene andere technische Systeme auf. Eine direkte Sicht z.B. auf Züge oder Gleise ist durch die örtliche Trennung der Arbeitsplätze von den überwachten Sektoren nicht mehr möglich.

Im Rahmen der Überwachung/Disposition erhalten Zugverkehrsleitende durch Systemhinweise Informationen zum Status von Zügen und über die Belegung von Strecken. Beispielsweise wird die Zugnummer von einem Zug besonders eingefärbt und unterstrichen dargestellt, wenn dieser vor einer Weiche stehenbleibt bzw. von einem System gestoppt wird. Entsprechend werden dem Zugverkehrsleitenden die Interpretation des Zustandes eines solchen Zuges und damit die Informationsverarbeitung durch das technische System erleichtert.

Bei der Behandlung von Störungen wird die Informationsverarbeitung durch die technischen Systeme nur wenig unterstützt. Es wird zwar angezeigt, dass ein bestimmtes Problem vorliegt, die Ursachen und Gründe muss der Zugverkehrsleitende jedoch auf der Basis seines Vorwissens und seiner individuellen Erfahrung erkunden.

Bei der Tätigkeit der Überwachung/Disposition wird die Entscheidungsfindung von den technischen Systemen kaum unterstützt. Im Rahmen der allgemeinen Betriebs- und Sicherheitsvorschriften müssen eigene Entscheidungen frei getroffen werden. Für die Tätigkeit der Störungsbehandlung ist die Entscheidungsfindung durch vordefinierte Prozesse (auch aus technischer Sicht) zum Teil stark eingeschränkt. Es existiert andererseits jedoch kein Expertensystem, welches den Zugverkehrsleitenden Entscheidungen diesbezüglich vollständig abnimmt.

Werkstatt-Diskussion

In einer nächsten Phase des Projekts sollen eine Abschätzung der künftigen technischen Entwicklungen und deren Einfluss auf die Tätigkeiten in der Zugverkehrsleitung stattfinden. Dazu werden im Rahmen von mehreren Workshops sowohl Entwickler wie auch Nutzer der technischen Systeme gemeinsam mögliche Zukunftsszenarien entwickeln und ihren Einfluss auf die verschiedenen Tätigkeiten in der Zugverkehrsleitung bewerten.

Im „Werkstatt Track“ soll diskutiert werden, mit welchen Methoden die Auswirkungen von Automatisierung auf ein soziotechnisches System am besten antizipiert und beurteilt werden können. Als Diskussionsgrundlage wird die Intuitive Logics Szenario-Methode vorgeschlagen (für einen Überblick siehe Kosow und Gaßner, 2008). Diese Methode gehört zu den kreativ-narrativen Szenariotechniken und bezieht neben objektiven Daten auch die Intuition und das implizite Wissen von Experten mit ein. Bezüglich der Kriterien, anhand derer die Auswirkungen der ausgearbeiteten Automatisierungsszenarios auf ein soziotechnisches System beurteilt werden sollen, schlagen die Autoren die drei von Manzey (2012) beschriebenen Problemfelder vor (Verlust von Situationsbewusstsein, Verlust von Fertigkeiten, übersteigertes oder zu geringes Vertrauen). Mit den Teilnehmenden des Werkstatt Tracks erhoffen sich die Autoren eine Diskussion der Vor- und Nachteile des Einsatzes der vorgeschlagenen Methoden und Kriterien sowie Vorschläge für die Umsetzung.

Des Weiteren soll gemeinsam diskutiert werden, ob Tätigkeiten mit spezifischen Automatisierungsprofilen auch mit spezifischen Risiken verbunden sind, und falls ja, mit welchen. Auf der Basis der bisherigen Projektergebnisse konnten bereits erste Hypothesen dazu aufgestellt werden. So wird z. B. vermutet, dass die Kombination eines mittleren Automatisierungsgrades bei der Informationsaufnahme und eines hohen Automatisierungsgrades in der Handlungsausführung (bei einer typischen Überwachungsaufgabe) verstärkt zu Monotonie und dadurch vermittelt zu einem geringeren Situationsbewusstsein führt.

Die Autoren erhoffen sich eine konstruktive, kritische Diskussion und gegebenenfalls methodischen Input von den Werkstattteilnehmenden.

Literatur

- Bainbridge, L. (1983). Ironies of Automation. *Automatica*, 19, 775-779.
- Endsley, M. R. (2000). Theoretical Underpinnings of Situation Awareness: A Critical Review. In M. R. Endsley & D. J. Garland (Hrsg.), *Situation Awareness Analysis and Measurement* (pp. 3-32). Lawrence Erlbaum Associates.
- Hauß, Y. & Timpe, K.-P. (2000). Automatisierung und Unterstützung im Mensch-Maschine-System. In K.-P. Timpe, T. Jürgensohn & H. Kolrep (Hrsg.), *Mensch-Maschine-Systemtechnik. Konzepte, Modellierung, Gestaltung, Evaluation*. Düsseldorf: Symposion (41-62).

- Kosow, H., & Gaßner, R. (2008). Methoden der Zukunfts- und Szenarioanalyse. Überblick, Bewertung und Auswahlkriterien. Werkstattbericht Nr. 103. Berlin: IZT - Institut für Zukunftsstudien und Technologiebewertung.
- Manzey, D. (2012). Systemgestaltung und Automatisierung. In P. Badke-Schaub, G. Hofinger & K. Lauche (Hrsg.), *Human Factors. Psychologie sicheren Handelns in Risikobranchen*. 2. überarbeitete Auflage (S. 333-352). Heidelberg: Springer.
- Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors*, 43, 563–572.
- Militello, L. G., & Hutton, R. J. (1998). Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. *Ergonomics*, 41(11), 1618-1641. Citeseer.
- Parasuraman R., Molloy R., & Singh I. L. (1993). Performance consequences of automation induced “complacency.” *International Journal of Aviation Psychology*, 2, 1–23.
- Parasuraman, R., & Riley, V., (1997). Humans and automation: use, misuse, disuse, abuse. *Human Factors* 39, 230-253.
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. (E. O. L. B. N. Laboratory, Ed.) *IEEE Transactions on Systems Man and Cybernetics Part A Systems and Humans*, 30(3), 286-297. IEEE.
- Wäfler, T., Windischer, A., Ryser, C., Weik, S. & Grote, G. (1999). Wie sich Mensch und Technik sinnvoll ergänzen. Die GESTALTUNG automatisierter Produktionssysteme mit KOMPASS. Schriftenreihe Mensch-Technik-Organisation (Hrsg. E. Ulich), Band 18. Zürich: vdf Hochschulverlag.

Erfassung psychomentaler Beanspruchung bei Arbeit mit unterschiedlichen kognitiven Anforderungen mit Hilfe des Elektroenzephalogramms (EEG)

Thea Radüntz¹ und Gabriele Freude¹

Schlüsselwörter: mental Workload, Beanspruchung, Elektroenzephalographie (EEG), Signalverarbeitung, Mustererkennung

Zusammenfassung

Im Mittelpunkt des Beitrages steht die Ermittlung psychomentaler Beanspruchung anhand einer aus unterschiedlichen kognitiven Anforderungen bestehenden – in der BAuA entwickelten – Aufgabenbatterie und mit Hilfe des Elektroenzephalogramms. 36 Probanden führten bis jetzt die kognitiven Aufgaben durch, weitere Experimente sind noch geplant. Die Probanden gaben subjektive Bewertungen zur Beanspruchung mittels NASA-TLX ab. Das EEG wurde während der Ausführung der Aufgaben mittels einer Elektrodenkappe registriert. Das Design der Signalverarbeitungs-Pipeline wird in dem Beitrag vorgestellt und die ersten Ergebnisse präsentiert. Diese belegen 1.) die Eignung der Aufgabenbatterie in Bezug auf die unterschiedliche Komplexität und Schwierigkeit der Aufgaben und der damit induzierten mentalen Beanspruchung und 2.) die Konsistenz der ersten Ergebnisse aus der EEG-Signalverarbeitung zu früheren Untersuchungen. Mit zunehmender Beanspruchung durch höheren Schwierigkeitsgrad sind eine Erhöhung des Theta-Anteils im frontalen Bereich und eine Abnahme des Alpha-Anteils im parietalen Bereich festzustellen.

Motivation und Zielstellung

Die Entwicklung moderner Informationstechnologien bis hin zu hoch interaktiven Arbeitsbereichen und Arbeitsassistenzsystemen ist unaufhaltsam. Trotz Vereinfachung der Arbeit klagen Arbeitnehmer jedoch über hohe Beanspruchung und Stress. Probleme ergeben sich durch Informationsflut, häufige Arbeitsunterbrechungen oder durch die Vielzahl irrelevanter Informationen (Kompier, 2001), (Landsbergis, 2003), (NIOSH, 2002).

Die Automatisierung kann aber auch mit monotonen Arbeitstätigkeiten verbunden sein, die den Aktivierungszustand der Beschäftigten herabsetzen (Parasuraman, 1993), (Parasuraman, 1994) und mit niedriger Wachsamkeit, erhöhter Ermüdung (Hacker, 1984), (Debitz, 2003), (May, 2008) und zunehmender Fehleranfälligkeit und damit einer Gefährdung weiterer Personen (Straeter, 2001) einhergehen.

¹ Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Gr. 3.4 "Mentale Gesundheit und kognitive Leistungsfähigkeit"

Die objektive Messung der Beanspruchung ist also dringend erforderlich, zumal eine langfristige Fehlbeanspruchung zu einem beträchtlichen Problem unserer modernen Gesellschaft werden kann.

Zielstellung der Untersuchungen ist es, fortlaufend und rückwirkungsarm die mentale Beanspruchung (im Weiteren auch als Workload bezeichnet) anhand ausgewählter EEG-Parameter zu erfassen. Der theoretische Hintergrund besteht in der Variabilität der EEG-Frequenzbänder in Abhängigkeit von Aufmerksamkeit und Ermüdung sowie der mentalen Beanspruchung.

Experimentelle Methoden

Die Simulation verschiedener Arbeitsanforderungen erfolgt mittels einer in E-Prime implementierten Aufgabenbatterie. Sie besteht aus Aufgaben unterschiedlicher Komplexität und Aufgabenschwierigkeit, die unterschiedliche Beanspruchung induzieren. Diese sind: 0-back (5 Minuten), 2-back (5 Minuten), Sternberg (10 Minuten), Sternberg mit serieller Buchstabenpräsentation (10 Minuten), Stroop (5 Minuten), Switch GER (5 Minuten), Switch NUM (5 Minuten), Switch XXX (10 Minuten) und AOSPAN (20 Minuten).

Die Untersuchung je Proband findet an einem Untersuchungstag statt. Sie besteht aus einer Trainingsphase und dem Hauptversuch. Während der Trainingsphase werden die Probanden mit den kognitiven Aufgaben des Experiments vertraut gemacht. Diese sind identisch mit den Aufgaben des Hauptversuchs, jedoch zeitlich deutlich kürzer. Sie werden so oft wiederholt, bis ein Genauigkeitsindex von mindestens 80% erreicht ist. Durch die Trainingsphase sollen vergleichbare Ausgangsbedingungen in Bezug auf die Leistung geschaffen und somit die Belastungseffekte im Hauptversuch unabhängig von Lerneffekten ermittelt werden. Nach jeder Trainingsaufgabe erfolgt eine computerbasierte Bewertung der NASA-TLX-Dimensionen durch den Probanden.

Der Hauptversuch findet nach einer kurzen Pause gleich im Anschluss an die Trainingsphase statt. Die Aufgaben werden in der gleichen ausbalancierten Reihenfolge wie in der Trainingsphase präsentiert, und nach jeder Aufgabe wird der NASA-TLX computerbasiert erhoben.

Die Stichprobe, die bis Ende 2013 untersucht werden soll, besteht voraussichtlich aus 100 Erwerbstätigen und lässt eine hohe Variabilität bzgl. der kognitiven Leistungsfähigkeit und Beanspruchung erwarten.

Die Versuche werden in der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin in einem abgeschirmten Labor durchgeführt. Die Steuerung des Experiments wird über eine Sprechanlage, eine Remote-Desktop-Verbindung und ein Videoüberwachungssystem gewährleistet.

Das EEG, als mittelbares Signal der menschlichen Informationsverarbeitung, wird, gegen Cz als Referenzelektrode, von 25 Elektroden nach dem internationalen 10/20-System mit einer Abtastfrequenz von 500 Hz während der Hauptunter-

suchung erhoben. Das vertikale Elektrokulogramm (EOG) wird an der IO1-Elektrode, das horizontale EOG an den Elektroden LO1 und LO2 abgeleitet. Für die EEG-Ableitung werden Verstärker und die Software „BrainRecorder“ der Firma „BrainProducts GmbH“ verwendet.

Weitere workload relevante physiologische Daten (z.B. HRV, BD) werden mit dem Finometer PRO der Firma MFS (Finapres Medical Systems) erhoben, damit eine Kombination bzw. Validierung anhand von verschiedenen Verfahren erfolgen kann.

Signalverarbeitung

Das System zur Berechnung des Workloads anhand vom EEG besteht aus verschiedenen Modulen, die als MATLAB-Toolbox implementiert sind. Diese sind in Abbildung 1 in Form einer Signalverarbeitungs-Pipeline dargestellt.

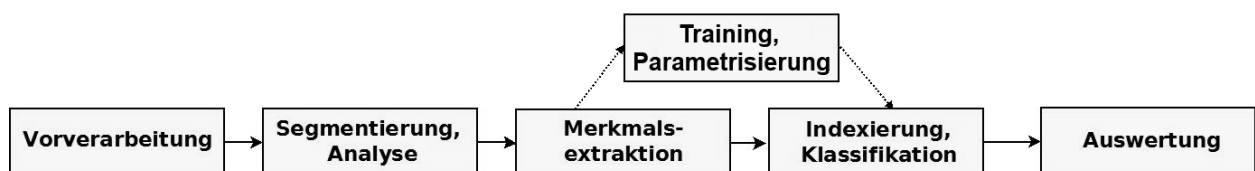


Abb. 1: EEG-Signalverarbeitungs-Pipeline.

Im Vorverarbeitungsmodul wird das EEG-Signal in MATLAB importiert und nach einer Hamming-Fensterung mit einem Bandpassfilter der Ordnung 100 zwischen den Grundfrequenzen 0.5 und 40 Hz gefiltert. Anschließend wird das Signal mittels ICA in seinen unabhängigen Komponenten zerlegt, welche vorerst visuell begutachtet, in Artefakt- bzw. EEG-Komponenten klassifiziert und für die weitere Verarbeitung entsprechend ausgewählt werden. Die EEG-Komponenten werden rückprojiziert und das artefaktbereinigte Signal zur Weiterverarbeitung dem nächsten Modul übergeben. Hier erfolgt eine Segmentierung in 10 Sekunden-Epochen mit einer Überlappung von 50% und die Transformation des EEG-Signals aus dem Zeit- in den Frequenzbereich mittels Fast Fourier Transformation. Das Leistungsspektrum und die relativen Frequenzbänder (θ mit 4 bis 8 Hz, α mit 8 bis 12 Hz) werden berechnet.

Das individuelle Training des Systems je Proband erfolgt, indem die Verteilungen der Bänder (θ und α) über alle Segmente einer Aufgabe berechnet und ihre Mittelwerte gespeichert werden. Das geschieht für eine Aufgabe je Schwierigkeitsstufe. Anschließend wird über alle für das Systemtraining ausgewählte Aufgaben und Segmente der Person die kumulative Verteilungsfunktion berechnet. Dann werden anhand der Mittelwerte je Schwierigkeitsstufe die entsprechenden p-Werte aus der kumulativen Verteilungsfunktion extrahiert. Diese werden zusammen mit den subjektiven NASA-Task-Load-Indizes für die individuelle Parametrisierung des Systems benötigt (siehe Lei, 2011).

Nach dem Training des Systems und der Generierung der individuellen Parameter b_0 , b_1 und b_2 erhalten wir für jeden Proband ein personalisiertes Logistic Func-

tion Model (LFM): $W = 1 / (1 + e^{-(b_0 + b_1 * \theta + b_2 * \alpha)})$. Hier können dann die relativen Werte der Frequenzbänder je Segment eingesetzt und ein Workload-Index W je Segment berechnet werden, der, wie man an den Eigenschaften der Funktion erkennen kann, im Wertebereich zwischen 0 und 1 liegt.

Erste Ergebnisse

Aufgrund der noch laufenden Untersuchungen beziehen sich die unten aufgeführten Ergebnisse auf die Daten der ersten 36 Probanden.

Erste Ergebnisse aus dem NASA-TLX sind in Abbildung 2 dargestellt. Sie zeigen, dass die geistige Dimension bei der konzipierten Aufgabenbatterie überwiegt. Damit kann davon ausgegangen werden, dass die Beanspruchung durch Informationsverarbeitungsprozesse induziert wird, welche sich im EEG widerspiegeln müssten. In Abbildung 2 (rechts) ist zu sehen, dass innerhalb der Aufgabenbatterie mehrere Schwierigkeitsstufen existieren. Damit sind ein erfolgreiches Training des Systems und die individuelle Parametrisierung gewährleistet. Da jede Schwierigkeitsstufe mehr als eine Aufgabe beinhaltet, existieren genug Daten für die Validierung des Systems, wo die Systemtrainingsaufgaben keine Verwendung finden sollten.

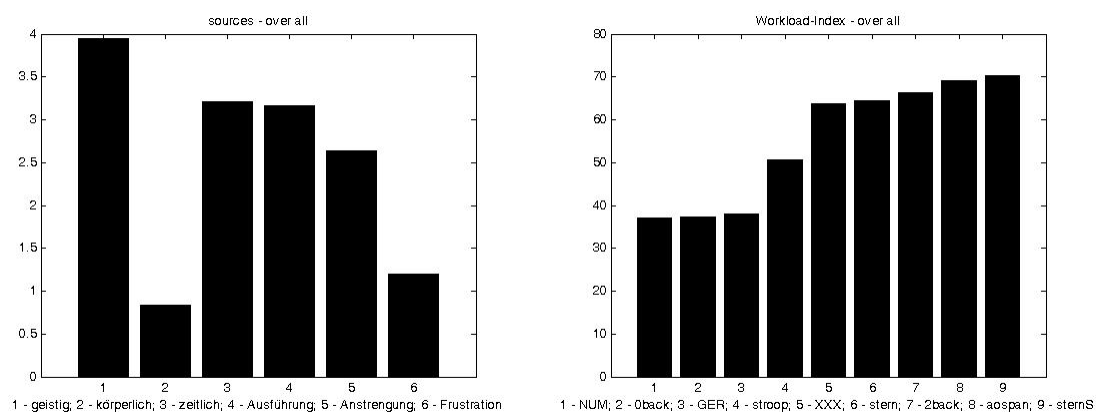


Abb. 2: links – Gewichtung der NASA-Dimensionen über die gesamte Aufgabenbatterie; rechts – NASA-TLX der Aufgaben über alle Probanden gemittelt

Die Frequenzbänder des EEG-Signals wurden nach einer Vorverarbeitung und Segmentierung vorläufig ausgewertet. Der Anteil der einzelnen Frequenzbänder bildet die Grundlage für die Indexierung des Workloads anhand des oben erwähnten LFM und lässt erste Tendenzen erkennen, die für die Güte des Modells von Bedeutung sein bzw. zu seiner Validierung und Verfeinerung beitragen könnten. Ausgewertet wurden vorerst die EEG-Daten der Aufgaben 0-Back, 2-Back und AOSPAN. Die Abbildung 3 zeigt die über alle Probanden und Segmente gemittelten relativen Leistungsspektren in den drei Aufgaben, gemessen an der frontalen Elektrode Fz (links) und an der parietalen Elektrode Pz (rechts). Dabei ist deutlich zu erkennen, dass mit erhöhtem Schwierigkeitsgrad der Anteil der Theta-Leistung insbesondere im frontalen Bereich (Fz) zunimmt und der Anteil der Alpha-

Leistung im parietalen Bereich (Pz) abnimmt. Abbildung 4 zeigt dieses Ergebnis in Form eines Balkendiagramms.

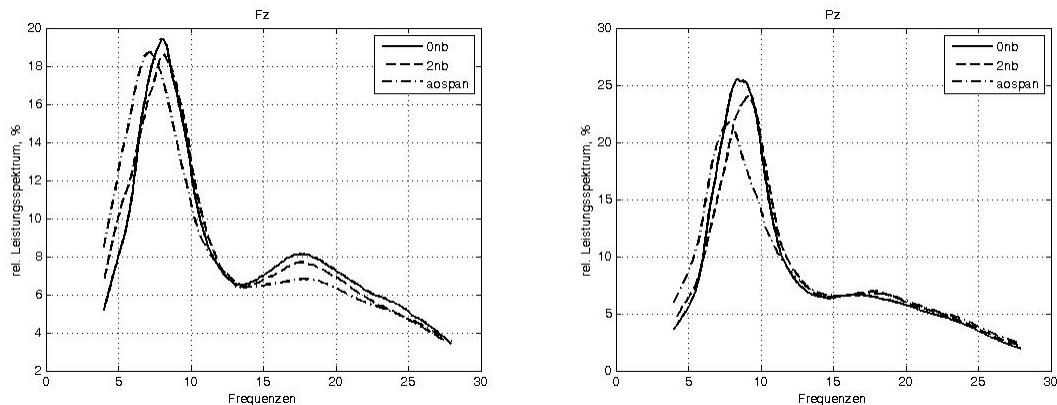


Abb. 3: links – relatives Leistungsspektrum an der Fz-Elektrode, rechts – relatives Leistungsspektrum an der Pz-Elektrode; gemittelt über alle Segmente und Probanden

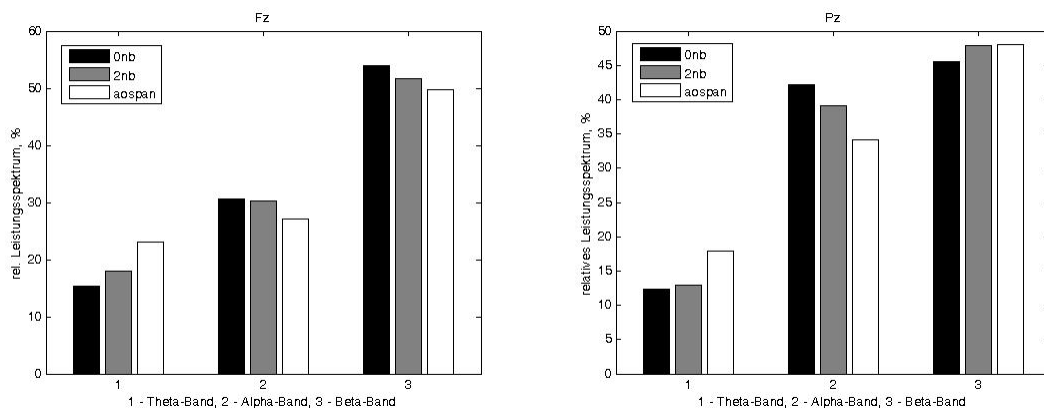


Abb. 4: links – Veränderung der Frequenzband-Anteile an der Fz-Elektrode, rechts – Veränderung der Frequenzband-Anteile an der Pz-Elektrode; je Aufgabe gemittelt über alle Segmente und Probanden

Die Ergebnisse stimmen mit den Beobachtungen in den Arbeiten von Lei, sowie mit Beobachtungen in anderen früheren Studien überein (z.B. Posner et al., 1990, 1992; Gundel & Wilson, 1992; Serman et al., 1994; Gevins et al., 1997, 1998; Wilson, 2001, 2002; Hagemann, 2008; Gevins and Smith, 2000; Smith et al., 2001; Sirevaag et al., 1988). Die Konsistenz dieser Tendenzen impliziert eine positive Aussagekraft des Modells und die individuelle Klassifikation in einem under load, normal load und overload Bereich (vergleich auch (Lei, 2011)).

Diskussion

Die Erfassung der mentalen Beanspruchung anhand ausgewählter EEG-Parameter stellt die zentrale Zielstellung der Studie dar. Dazu wurde eine Aufgabenbatterie entwickelt, die aus Aufgaben unterschiedlicher Komplexität und Aufgabenschwierigkeit besteht und damit verschiedene Arbeitsanforderungen simuliert sowie unterschiedliche Beanspruchung induziert. Die Ergebnisse der ersten 36 subjektiven

Bewertungen der Probanden anhand des NASA-TLX belegen die Richtigkeit dieser Annahmen und zeigen, dass die Beanspruchung überwiegend auf die kognitive Informationsverarbeitung zurückzuführen ist.

Wie in zahlreichen Untersuchungen beobachtet, stellt das EEG als mittelbares Signal der menschlichen Informationsverarbeitung bzw. die Variabilität seiner Frequenzbänder die zentrale Annahme der Studie bezüglich einer objektiven Methode zur Registrierung von mentaler Beanspruchung dar. Die Konsistenz der hier erzielten ersten Ergebnisse zu früheren Studien, lassen die Richtigkeit der Annahme vermuten.

Nächste Schritte zielen auf die individuelle Parametrisierung des Systems und auf die anschließende Validierung des Logistic Function Models ab. Außerdem soll die momentan noch visuelle Klassifikation der unabhängigen Komponenten aus dem EEG-Signal in Artefakt- und EEG-Komponenten automatisiert werden. Dazu werden verschiedene Methoden des maschinellen Lernens Anwendung finden (Neuronale Netze, Support Vector Machines, etc.). Des Weiteren wird überprüft, ob die Verwendung der Wavelet-Transformation für die Merkmalsextraktion mehr Information aus dem EEG-Signal in Bezug auf die Bänderleistungen schöpft und ob es gewinnbringend ist, die Komponenten der ICA (Independent Component Analysis) direkt in die Merkmalsextraktion einfließen zu lassen.

Aus den Ergebnissen kann schließlich ein Prototyp resultieren. Dieser soll aus mobiler Sensorik und einer echtzeitfähigen Implementierung des Verfahrens bestehen. Er kann entweder auf Mikroprozessor-Ebene oder als integrierter Schaltkreis realisiert werden. Sein Einsatz als Assistenzsystem bei der täglichen Arbeit kann den Arbeitnehmer vor Überlastung oder Unterforderung warnen. Vorstellbar ist auch, das Verfahren zur Bewertung informationsintensiver Tätigkeiten zu benutzen, um neuartige Mensch-Maschine- Systeme anhand des entstehenden Workloads zu validieren.

Literatur

- Debitz, U., Gruber, H., & Richter, G. (2003). Psychische Gesundheit am Arbeitsplatz. Teil 2: Erkennen, Beurteilen und Verhüten von Fehlbeanspruchungen (3rd ed.), InfoMediaVerlag.
- Gevins, A., Smith, M.E., McEvoy, L., Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: effects of difficulty, type of processing, and practice. *Cerebral Cortex*, 7, 374-385.
- Gevins, A., Smith, M.E., Leong, H., McEvoy, L., Whitfield, S., & Du, R. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors*, 40, 79-91.
- Gevins, A., & Smith, M.E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex*, 10(9), 829-839.

- Gundel, A., & Wilson, G.F. (1992). Topographical changes in the ongoing EEG related to the difficulty of mental tasks. *Brain Topography*, 5, 17-25.
- Hacker, W., & Richter, P. (1984). Psychische Fehlbeanspruchung. Psychische Ermüdung, Monotonie, Sättigung und Stress (Spezielle Arbeits- und Ingenieurpsychologie in Einzeldarstellungen) (2nd ed.), Berlin: Springer.
- Hagemann, K. (2008). The alpha band as an electrophysiological indicator for internalized attention and high mental workload in real traffic driving. Dissertation, Universität Düsseldorf, Deutschland.
- Kompier, M. A. J., & Kristensen, T. S. (2001). Organisational work stress interventions in a theoretical, methodological and practical context. In J. Dunham (Eds.), *Stress in the workplace: past, present and future* (pp. 164-190). London: Whurr Publishers.
- Landsbergis, P. A., Cahill, J., & Schnall, P. (2003, January). The changing organisation of work and the safety and health of working people: a commentary. *Journal of Occupational Environmental Medicine*, 45(1), 61-72.
- Lei, S. (2011). Driver mental states monitoring based on brain signals. Retrieved from http://opus.kobv.de/tuberlin/volltexte/2011/3185/pdf/lei_shengguang.pdf
- May, J. F., & Baldwin, C. L. (2008, November). Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transportation Research, Part F* 12 (2009), 218-224.
- NIOSH - NORA Organization of work team members (2002, April). *The changing organization of work and the safety and health of working people*. Cincinnati: NIOSH-Publications Dissemination.
- Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation induced complacency. *International Journal of Aviation Psychology*, 3(1), 1-23.
- Parasuraman, R., Mouloua, M., & Molloy, R. (1994). Monitoring automation failures in human machine systems. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: current research trends* (pp. 45-49). Hillsdale, NJ: Earlbaum.
- Posner, M.E., & Peterson, S.E. (1990). The attentional system of the human brain. *Annual Review of Neuroscience*, 13, 25-42.
- Posner, M.E., & Rothbart, M.K. (1992). Attentional mechanisms and conscious experience. In: A.D. Milner & M.D. Rugg (Eds). *The neuropsychology of consciousness*. San Diego: Academic, 91-111.
- Sirevaag, E., Kramer, A.F., De Jong, R., & Mecklinger, A. (1998). A psychophysiological analysis of multi-task processing demands. *Psychophysiology*, 25, 482.
- Smith, M.E., Gevins, A., Brown, H., Karnik, A., & Du, R. (2001). Monitoring task load with multivariate EEG measures during complex forms of human-computer interaction. *Human Factor*, 43(3), 366-380.

- Sterman, M.B., Mann, C.A., Kaiser, D.A., & Suyenobu, B.Y. (1994). Multiband topographic analysis of a simulated visuomotor aviation task. *International Journal of Psychophysiology*, 16, 49-56.
- Sträter, O. (2001). Warum passieren menschliche Fehler und was kann man dagegen tun?, Forum Prävention, AUVA - Allgemeine Unfallversicherungsanstalt, Wien. Retrieved from http://www.eval.at/fp11/pdf/Plenar_Mittwoch/Straeter.pdf
- Wilson, G.F. (2001). In-flight psychophysiological monitoring. In F. Fahrenberg & M. Myrtek (Eds.), *Progress in ambulatory monitoring* (pp. 435-454). Seattle, WA: Hogrefe and Huber.
- Wilson, G.F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology*, 12(1), 3-18.

Eye Tick - Blickinteraktion im öffentlichen Raum

Antje Venjakob¹, Juliane Bock¹, Florian Hoppe¹, Peter Brandstädter¹, Stefan Ruff¹

Schlüsselwörter: Eye Tracking, Blickinteraktion, Verweildauer, Blinzeln, Blickgesten

Zusammenfassung

Das Projekt 'Eye Tick' untersucht blickbasierte Interaktionsformen für den öffentlichen Raum. In zwei Experimenten wurden Rahmenbedingungen für diese am Beispiel eines Fahrkartenautomaten untersucht. Da gängige Eye Tracker mit einer Kalibrierung einhergehen, wurde anhand einer statischen Anbringung des Eye Trackers untersucht, ob eine 2-Punkt Kalibrierung bei variierender Körpergröße und dem Vorhandensein einer Brille ausreichende Genauigkeit für eine erfolgreiche Bedienung bietet. Dies konnte bejaht werden. Zudem wurde eine Umsetzung des Konzepts der Blickgesten mit dem Ziel getestet, das sogenannte Midas Touch Problem zu umgehen. Bei der hier umgesetzten Form der Blickgesten wurde für diese Interaktionsform signifikant mehr Zeit benötigt, als für eine Interaktion mittels Verweildauer oder Blinzeln. Deskriptiv zeigte sich jedoch auch, dass die Interaktion mittels Blickgesten weniger fehleranfällig ist als die anderen beiden Konzepte. In weiteren Experimenten wird der Frage nachgegangen, wie Blickinteraktion gestaltet sein sollte, um eine sichere und zuverlässige Interaktion im öffentlichen Raum durch unterschiedliche Nutzergruppen zu ermöglichen.

Motivation

Die Gestaltung von neuen Interaktionskonzepten, die sich von der klassischen Maus-/Tastatureingabe entfernen, ist eines der Kernthemen der neueren Mensch-Maschine Forschung. Die Eingabe über den Blick erscheint hierbei besonders attraktiv, weil sich die Aufmerksamkeit des Nutzers und der Blickort überschneiden. Verschiedene Experimente haben gezeigt, dass dies Geschwindigkeitsvorteile bei der Bedienung von Oberflächen bietet (Sibert & Jacob, 2000). Die breite Anwendung von Blickinteraktion birgt jedoch auch Probleme, wie das Midas Touch Problem (Jacob, 1990), welches sich darauf bezieht, dass bei der Interaktion über Verweildauer, alle Objekte, die visuelle Aufmerksamkeit auf sich ziehen ausgewählt werden.

Das Projekt 'Eye Tick' hat es sich zur Aufgabe gemacht, zu untersuchen, wie entsprechende Hindernisse der Blickinteraktion überwunden werden können. Die zentrale Idee hierbei ist weniger die Nutzung von Blickinteraktion als personalisierte Eingabemöglichkeit, z.B. am Computer, sondern die Übertragung von Blick-

¹ Technische Universität Berlin, Institut für Psychologie und Arbeitswissenschaft, Fachgebiet Mensch-Maschine Systeme

interaktion in den öffentlichen Raum. Die Herausforderung hierbei ist es den Kommunikationsweg so zu gestalten, dass er bei einer größtmöglichen Nutzergruppe mit geringstmöglichen Fehlbedienungen funktioniert. Als Fallbeispiel wurde die blickbasierte Interaktion mit einem Fahrkartenautomaten gewählt. Um in diesem Kontext eine möglichst reibungslose Bedienung zu ermöglichen, wird ein ganzheitlicher Ansatz verfolgt:

Da die Implementierung einer Blicksteuerung im öffentlichen Raum am einfachsten mit einem statischen, fest montierten Eye Tracker zu realisieren ist, wurde untersucht, welchen Einfluss anthropometrische Eigenschaften auf die Kalibrierungsgüte haben (Experiment 1). In einem zweiten Experiment wurde geprüft, ob neben den klassischen Auswahlmöglichkeiten Verweildauer und Blinzeln (Koesling, Zöllner, Sichelschmidt & Ritter, 2009) noch andere Interaktionskonzepte in Frage kommen. Hierzu wurde das Konzept der Blickgesten (Drewes & Schmidt, 2007) vergleichend gegen die beiden klassischen Interaktionsformen getestet (Experiment 2).

Experiment 1

Ziel des ersten Experiments ist es herauszufinden, welche Kalibrierungsgüte mit einem statisch angebrachten Eye Tracker für verschiedene Personengruppen erreicht werden kann. Zu diesem Zweck wurde die Kalibrierungsgüte bezüglich verschiedener Körperlängen und dem Vorhandensein einer Brille erhoben. Es wird hierbei erwartet, dass sich eine große bzw. kleine Körperlänge auf Grund der statischen Anbringung des Eye Trackers negativ auf die Kalibrierung auswirken würden im Vergleich zu einer mittleren Körperlänge. Ebenso wird erwartet, dass sich das Vorhandensein einer Brille wegen möglicher optischer Effekte unvorteilhaft auf die Kalibrierungsgüte auswirkt.

Methode

In einem Vortest mit neun Versuchsteilnehmern wurde ein Winkel von 128° zwischen mobilem Eye Trackinggerät (SMI RED-oem, Sampling Rate 60 Hz) und Monitor bei einem Abstand der Versuchsperson zwischen 35 und 103 cm zum Monitor (Eizo FlexScan L367, 15“ LCD Display mit einer 1024 x 768 Vollbildauflösung) als optimal bestimmt. Monitor und Eye Tracker waren in das Mock-up eines Fahrkartenautomaten eingebunden, das aus einer Pappummantelung besteht, die den tatsächlichen Abmessungen eines Höft & Wessel AG Automaten entspricht, wie er zum Beispiel von der Deutschen Bahn verwendet wird (Abb.1).

Für den Versuch wurde eine 2-Punkt Kalibrierung gewählt, da für die Anwendung im öffentlichen Raum mehr Kalibrierpunkte als zu zeitaufwendig erscheinen und dem Ziel einer effizienten Interaktion entgegen wirken. Mit jeder Versuchsperson wurden vier Kalibrierungen mit jeweils einer 2-Punkt Validierung durchgeführt. Zwischen den einzelnen Kalibrierungen wurden die Versuchspersonen gebe-

ten vom Eye Tracker wegzutreten und einen Fragebogen auszufüllen. Dies diente vornehmlich dazu, dass sich die Versuchsteilnehmer vor jeder Kalibrierung neu vor dem Eye Tracker ausrichten.

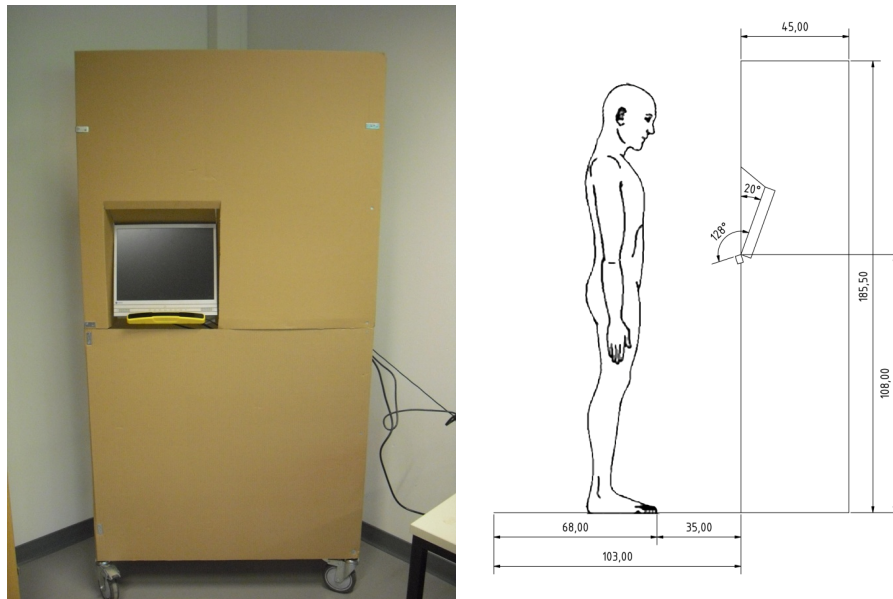


Abb.1:Mock-Up des verwendeten Fahrkartenautomaten.

Es wurden zwei between-subjects Faktoren erhoben: Der Faktor 'Körperlänge' ist dreifach gestuft und verfügt über die Faktorstufen 'klein'(153-162 cm), 'mittel' (163-175 cm) und 'groß' (176-187 cm). Diese Maße leiten sich aus der DIN 33402-2:2005-12 her und umfassen 90% der deutschen Bevölkerung zwischen 18 und 65 Jahren. Der Faktor 'Vorhandensein einer Brille' gliedert sich in eine Versuchspersonengruppe mit und eine ohne Brille. Als abhängige Variable wurde die Abweichung in x und y Richtung vom tatsächlichen Ort der Validierungspunkte erhoben. Insgesamt nahmen 24 Versuchspersonen mit einem Durchschnittsalter von 26 Jahren ($SD= 2,7$; Range= 20-32 Jahre) an der Untersuchung teil. Die Versuchspersonen verteilten sich gleichmäßig über die Experimentalbedingungen, so dass jede Faktorstufenkombination von vier Versuchspersonen ausgeführt wurde.

Resultate

Die Inspektion der Ergebnisse zeigt zunächst nur kleinere Schwankungen der Kalibriergüte, wie Tabelle 1 entnommen werden kann. Auf Grund der sehr kleinen Stichprobengröße pro Zelle wurden die Daten mittels des nicht-parametrischen Kruskal-Wallis Tests für die beiden Faktoren getrennt ausgewertet. Hierbei ergab sich ein signifikanter Unterschied bezüglich der y-Werte des Faktors 'Körperlänge' ($U=6,417$, $p=0,04$). Ein paarweiser Vergleich der Faktorstufen führte den Effekt auf Unterschiede bezüglich der Faktorstufen 'mittel' und 'groß' zurück.

Tab.1: Mittlere Kalibriergüte in ° Blickwinkel, Standardabweichungen in Klammern. x Werte links, y Werte rechts

	Klein		Mittel		Groß	
Brille	0,49 (0,06)	0,58 (0,04)	0,49 (0,05)	0,44 (0,03)	0,49 (0,14)	0,55 (0,05)
Keine Brille	0,41 (0,12)	0,47 (0,14)	0,38 (0,06)	0,45 (0,07)	0,48 (0,17)	0,56 (0,17)

Diskussion

Die Ergebnisse dieses Experiments weisen trotz der starren, einheitlichen Anbringung des Eye Trackers gute Ergebnisse für alle Versuchsbedingungen auf. Die Mittelwerte aller Kategorien blieben in x-Richtung unter der häufig verwendeten Grenze akzeptabler Kalibrierungen von $0,5^\circ$ Blickwinkel, während die Mittelwerte in y-Richtung diese in vielen Fällen knapp überschritten. Dies ist in Einklang mit Befunden, wonach Blickbewegungen in y-Richtung ungenauer sind als in x-Richtung (Collewyn, Erkelens & Steinman, 1988). Interessanterweise, unterscheidet sich die Kalibriergüte lediglich hinsichtlich der y-Ausprägung zwischen Teilnehmern mit mittleren und großen Körpergrößen. Auf Basis dieses Experiments kann also geschlussfolgert werden, dass eine starre Anbringung eines Eye Trackers im öffentlichen Raum möglich ist, ohne dass große Gruppen der Bevölkerung von der Nutzung prinzipiell ausgeschlossen wären. In diesem Rahmen muss allerdings darauf hingewiesen werden, dass die Kategorien der Körperlängen keine Extremwerte umfassten. Die größte Abweichung insgesamt betrug sowohl in x- wie auch in y-Richtung $0,7^\circ$ Blickwinkel. Bei einem mittleren Abstand zum Monitor von 60 cm entspricht dies einer Abweichung von 0,73 cm. Bei der derzeitigen Größe der Schaltflächen von S-Bahnautomaten in Berlin (9,5cm in x Richtung und 1,9 cm in y-Richtung) wäre bei einem ungefähr mittigen Blick eine Bedienung selbst bei relativ schlechter Kalibriergüte noch möglich.

Experiment 2

Das Midas Touch Problem ist besonders mit der Auswahl von Objekten über die Verweildauer verbunden, da jede verlängerte Betrachtung automatisch zu einer Auswahl des jeweiligen Objektes führt. Ziel dieses Experimentes war es Blickgesten als eine zuverlässigere Form der blickgestützten Kommunikation im Vergleich zur Methode der Verweildauer und des Blinzeln zu testen. Die in diesem Experiment verwendete Blickgeste wurde in einem Vorexperiment mit 12 Probanden aus sechs verschiedenen getesteten Blickgesten aufgrund ihrer Robustheit, der relativ schnellen Durchführbarkeit und der subjektiven Bewertung der Teilnehmer ausgewählt.

Methode

Für das vorliegende Experiment wurde die Oberfläche eines Fahrkartenautomaten der Berliner S-Bahn in „Microsoft Visual Studio 2010“ nachgebildet. Die Nachbildung umfasst Anzahl und Größe der Schaltflächen sowie Kaufoptionen, Gliederungsebenen und Farben. Jede der 12 Versuchspersonen (Alter: $M = 28,25$ Jahre, $SD = 3,29$ Jahre) führte vier Aufgaben mit drei Interaktionsformen aus. Jede dieser insgesamt 12 Aufgaben war durch die Betätigung von drei Schaltflächen lösbar. Das Ziel war es jeweils ein S-Bahnticket zu kaufen. Die Oberfläche wurde auf einem 17" LCD Monitor mit einer Bildschirmauflösung von 1024 x 768 Pixeln und einer Bildschirmaktualisierungsrate von 60 Hertz dargeboten.

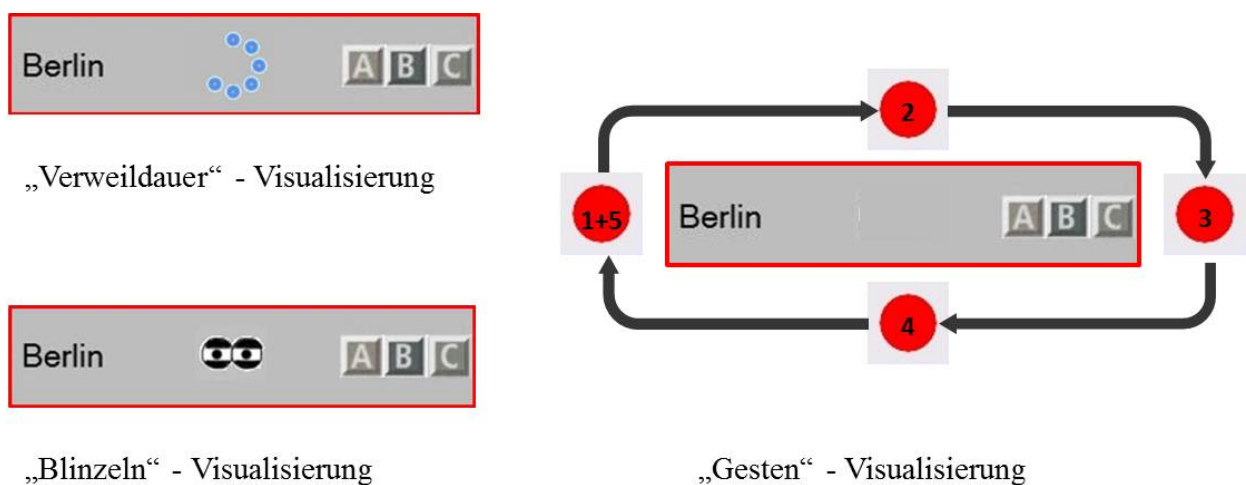


Abb.2: Visualisierung der Interaktionsform

Die Interaktionsform diente in diesem einfaktoriellen within-subjects Versuchsdesign mit drei Faktorstufen als unabhängige Variable. Befindet sich der Blick bei der Faktorstufe 'Verweildauer' an nicht-aktiven Stellen des Monitors, so ist der Blickzeiger ausgeblendet. Befindet sich der Blick auf einem aktiven Button, deuten Punkte einen sich aufbauenden Kreis an (siehe Abb.2). Nach 400 ms Verweildauer ist dieser Kreis geschlossen und der Button wird ausgewählt. Beim Blinzeln wird ein sich schließendes Augenpaar eingeblendet, wenn sich der Blick über einem Button befindet. Werden über eine Dauer von mehr als 400 ms keine Blickdaten aufgezeichnet, wird die Schaltfläche, auf der sich die letzte Fixation befand, ausgewählt. Bei den Blickgesten wird zunächst eine Schaltfläche über eine Verweildauer von 400 ms ausgewählt. Daraufhin startet die Geste indem ein roter Hinweispunkt links der Schaltfläche erscheint. Sobald ein Blicksample in einen quadratischen Bereich von 56 x 56 px um diesen Punkt fällt, verschwindet dieser und erscheint an anderer Stelle der Schaltfläche (siehe Abb.2), bis die Blickgeste einen Kreis schließt. Während der Blick von dem Punkt um die Schaltfläche herum geleitet wird sind alle Schaltflächen für 1,5 Sekunden pro eingeblendeten Hinweispunkt inaktiv, so dass bei der Ausführung der Geste keine anderen Schaltflächen aktiviert werden. Fällt innerhalb von 1,5 Sekunden kein Blicksample in den Be-

reich eines Hinweispunkts, so wird die Geste abgebrochen. Als abhängige Variablen dieses Experimentes dienen die durchschnittliche Bearbeitungszeit der Aufgaben sowie die Anzahl der Fehlbedienungen beim Kauf der Fahrkarten.

Resultate

Die Ergebnisse des Experiments zeigen, dass die Interaktionsform Verweildauer die schnellste Aufgabenerfüllung ermöglicht, gefolgt von der Interaktion über Blinzeln. Für die Interaktion über Gesten wurde die meiste Zeit benötigt, siehe Tabelle 2. Eine einfaktorielle ANOVA bestätigte, dass es sich hier um einen signifikanten Unterschied handelt ($F(2,22)=16,1$, $p<0,001$). Ein Bonferroni Vergleich der einzelnen Faktorstufen ergab, dass sich die Werte aller drei Bedingungen signifikant voneinander unterscheiden.

Tab.2: Mittlere Bediendauer der einzelnen Interaktionsformen in Sekunden, Standardabweichung in Klammern

	Verweildauer	Blinzeln	Blickgesten
M	9,45	18,80	36,04
(SD)	(4,29)	(7,95)	(16,54)

Der Vergleich der Anzahl von Fehlbedienungen ergab keinen signifikanten Unterschied. Die Mittelwerte und Standardabweichungen können Tabelle 3 entnommen werden.

Tab.3: Mittlere Anzahl Fehlbedienungen pro Interaktionsform, Standardabweichung in Klammern

	Verweildauer	Blinzeln	Blickgesten
M	1,21	0,79	0,35
(SD)	(1,52)	(0,65)	(0,41)

Diskussion

Die Ergebnisse zeigen, dass die Interaktion mittels Verweildauer am schnellsten funktioniert. Dies ist nicht überraschend, weil diese Interaktion in ihrer physischen Ausführbarkeit die schnellste ist. Interessanterweise zeigt sich in der deskriptiven Statistik aber auch, dass diese Interaktionsform am fehleranfälligsten ist und am meisten schwankt. Bezüglich der Robustheit gegenüber unwillentlichen Eingaben zeigt sich hier ein Vorteil für die Blickgesten. Es ist jedoch zu betonen, dass diese Ergebnisse nicht statistisch gesichert sind. Hier empfiehlt es sich, ein ähnliches Experiment mit einer größeren Anzahl von Aufgaben pro Modalität sowie einer größeren Teilnehmerzahl durchzuführen.

Ausblick

Das Projekt 'Eye Tick' ist mit den zwei beschriebenen Experimenten nicht zu Ende. In einem weiteren Experiment wurde die Gestaltung des Displays in Abhängigkeit von der Interaktionsform in den Fokus gerückt. Hierzu wurden die zwei klassischen Auswahlfunktionen der Blicksteuerung, die Verweildauer auf einem Objekt und verlängertes Blinzeln hinsichtlich ihrer Eignung für die Steuerung dreier auf die jeweiligen Interaktionsformen optimierter Oberflächen geprüft.

Zudem finden weitere Experimente statt, die die Eignung von feedbackloser Blicksteuerung für die Eingabe eines PIN-Codes zur Zahlung der Fahrkarte untersuchen. Hierbei steht die sichere und unbeobachtbare Eingabe der PIN im Fokus.

Acknowledgement

Für die möglichst anwendungsnahe Umsetzung der Versuche wurde das Projekt 'Eye Tick' bei der Anschaffung eines RED-oem Eye Trackers von Sensomotoric Instruments unterstützt. Dafür möchten wir uns herzlich bedanken.

Literatur

- Collewijn, H., Erkelens, C.J. & Steinman, R.M. (1988). Binocular co-ordination of human vertical saccadic eye movements. *Journal of Physiology*, 404, 183-197.
- Drewes, H & Schmidt, A. (2007). Interacting with the Computer using Gaze Gestures. *Proceedings of the 11th International Conference on Human-Computer Interaction (INTERACT 2007)*, S. 457-488. Berlin: Springer, 2007.
- Jacob, R.J.K. (1990). What you look at is what you get: eye movement-based interaction techniques. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, S.11-18, Seattle: ACM.
- Koesling, H., Zöllner, M., Sichelschmidt, L., & Ritter, H. (2009). With a flick of the eye: Assessing gaze-controlled human-computer interaction. *Cognitive Systems Monographs, Human Centered Robot Systems: Cognition, Interaction, Technology*. S. 83–92. Berlin: Springer.
- Sibert, L.E. & Jacob R.J.K. (2000). Evaluation of eye gaze interaction, *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 281-288, New York: ACM.

A Light-Weight Mixed Method Tool for Remote UX Evaluation: A Case Study of Premium Sports Watch Users

Piia Nurkka¹ and Satu Jumisko-Pyykkö²

Keywords: user experience, evaluation, statements, sentence completion, research method

Abstract

The goal of this paper was to develop a light-weight mixed method for assessment of remote online user experience and to evaluate it. The method was iteratively developed and is composed of quantitative section with statements and qualitative section with sentence completion task. A case study was conducted to test the tool. Data was collected remotely from 35 users of premium sports watch. The different methods produced consistent and complementary results providing a more comprehensive and reliable picture of the user experience. The method can be used by practitioners and academia for studying user experience.

Evaluation of User Experience

New evaluation methods are needed for understanding user experience (UX). UX is constructed in interaction with technology and is subjective, situated, complex and dynamic in its nature (Hassenzahl & Tractinsky, 2006). It is influenced by a user, system and the context of use (ibid). Experience as an outcome of the product perception is formulated on pragmatic and hedonic dimensions (Hassenzahl, 2003). The pragmatic attributes are connected to goal achievement and described by utility and usability (ibid). Hedonic attributes characterizes user's self and can be subdivided into stimulation (e.g. challenge and surprise) and identification (e.g. self-expression and personal values) (ibid). To measure UX, multiple methods has been proposed in the past, but there is need for better understanding of characteristics, procedures, benefits, weaknesses and validity of these methods and, thus, space for further development (see review Bargas-Avila & Hornbæk, 2011). For example, the practitioners value cost-effective, lightweight and easy to use tools capable to reach users on global markets remotely when assessing UX as a part of product development (Vermeeren et al., 2010; Väänänen-Vainio-Mattila et al., 2008). For academy, well-validated and reliable tools are needed for constructing deeper understanding on UX.

The major characteristics of traditional quantitative research are a focus on deduction, confirmation, theory/hypothesis testing, explanation, prediction, standardized data collection, and statistical analysis (Johnson & Onwuegbuzie, 2004).

¹ Tampere University of Technology

² Tampere University of Technology

Questionnaire is the most typical quantitative UX research method (used by 44-53% of the studies) (Bargas-Avila & Hornbæk, 2011; Vermeeren et al., 2010). The granularity of numerous existing UX questionnaires varies from assessment of overall experience to its specific components. AttrakDiff¹, based on pragmatic/hedonic model, measures Perceived Pragmatic Quality and Perceived Hedonic Quality for evaluating constructs of Beauty and Goodness with a total of 21 items on a 7-point bipolar scale with semantic differentials (Hassenzahl, 2004). There are also questionnaires to evaluate specific components of UX such as aesthetics (Lavie & Tractinsky, 2004) or emotion (Lang, 1980). The newly developed UX methods are rarely throughout validated and have limited content validity (in relation to UX models) (Bargas-Avila & Hornbæk, 2011). For practitioners, questionnaires are useful lightweight tools for understanding user's needs and requirements, benchmarking and improving products (Root & Draper, 1983). However, they are limited in providing explanatory information or reasoning behind the assessment (e.g. meanings).

Traditional qualitative research methods are characterized by induction, discovery, exploration, theory/hypothesis generation, the researcher as the primary instrument of data collection, and qualitative analysis (Johnson & Onwuegbuzie, 2004). In HCI community, the use of qualitative methods has increased in their popularity over time, and currently approximately half of UX evaluation research conducted is qualitative (Bargas-Avila & Hornbæk, 2011, Vermeeren et al., 2010, Barkhuus & Rode, 2007). Qualitative research, highlighting details and richness, is valued for constructing understanding on UX and its themes, such as aesthetics, affect and meaning of technology for users (Bargas-Avila & Hornbæk, 2011; Vermeeren et al., 2010). Parallel to conventional qualitative techniques (e.g. interviews, focus groups, observations) projective and constructive UX research methods have become more popular (Bargas-Avila & Hornbæk, 2011). In projective technique, the person is asked to make sense of the ambiguous stimulus by having to add to it (e.g. by filling out a speech bubble or a sentence) (Soley & Smith, 2008). By filling out the missing part, the subject projects part of him into it, and hence information is obtained about the person (ibid). Projective techniques, like association, completion, construction, choice or ordering types, have been used in clinical psychology (to assess e.g. personality and motivation), marketing research (consumer attitudes), and UX research (user values and meaning, UX evaluation) (Soley & Smiths, 2008; Nurkka et al., 2009). While qualitative methods can extend understanding on UX, the data-collection and analysis can be laborious and require special knowledge for identifying UX factors (e.g. identification, Vermeeren et al., 2010) for practitioners.

In remote studies, evaluators are separated from users in space and/or time, and the data-collection is conducted online over the Internet [adapted from Hartson et al., 1996]. The method is used for reaching cost-effectiveness (time, resources) and

¹ <http://www.attrakdiff.de/en/AttrakDiff/What-is-AttrakDiff/>

for global sampling (Bruun et al., 2009; Vermeeren et al., 2010). Among numerous (38) UX methods, questionnaire is the most typical remote online evaluation tool while a few qualitative methods have also been listed (diary, online interview, sentence completion) (see review Vermeeren et al., 2010). The methods are rather easy to implement to online environment, however, issues in validity (cross-cultural sampling, drop-out rates) and reliability are rarely addressed (Evans & Mathur, 2008).

Mixed method research combines quantitative and qualitative research techniques, methods, approaches, concepts, or language into a single study being suitable for applied research (Tashakkori & Teddlie, 2008). They are used to provide complementary viewpoints, to provide a complete picture of phenomena, to expand the understanding of phenomena, and to compensate for the weaknesses of one method (ibid). As technology gets more and more complex, exploration and use of various combinations of emerging and conventional methods is anticipated to enable gathering new knowledge, to build complete image of phenomenon and establish credibility for the findings (Barkhuus & Rode, 2007; Hagen et al., 2005; Law & van Schaik, 2010). Though one third of recent UX studies report collecting both quantitative and qualitative data, the purpose and procedure (e.g. analysis) of mixed method are left unclear (Vermeeren et al., 2010).

The goal of this paper is to develop and evaluate a mixed method tool for assessment of UX remotely. We construct the tool utilizing questionnaire and sentence completion techniques. We evaluate the method using meta-analysis to investigate the compatibility of the techniques by identifying divergent and convergent results, and to guide further development of the method presented.

Tool development

There were five goals in the UX evaluation tool development: 1) to aim at good content validity by basing the tool on UX theory and existing methods, 2) to provide both reliable quantitative and qualitative results to support product development, 3) to have a light-weight data-collection and analysis procedure (easy and cost-effective to use for practitioners), 4) to enable remote data-collection for reaching users worldwide, 5) to have good internal validity (easy, understandable and short to fill for users). The method was developed in collaboration with a manufacturer of sports instruments. A Light-Weight Mixed Method Tool for Remote UX Evaluation is composed of two parts; questionnaire with statements and sentence completion tasks.

The questionnaire development procedure composed item development and refinement. In the development of questionnaire, four UX dimensions were extracted from the model (Hassenzahl, 2003), operationalized, and related measurement instruments (statements) were defined (see Figure 1). In generating the statements, existing questionnaires from marketing (Tsai, 2005; Strizhakova et al., 2008) and design (Schifferstein & Zwartkruis-Pelgrim, 2008), reflecting essential aspects of

UX, were used and modified. A pool of 47 candidate statements was collected and cross-checked against the model (Hassenzahl, 2003). Then, the pool of statements was reviewed by five UX experts (academy/industry) and a naïve pilot evaluator to improve validity/reliability of the tool. The review resulted in deletion of duplicate statements, rephrasing of some statements, adding new statements in Utility dimension and creation of new dimension "Overall judgment" with three statements. At the end, there were 24 statements (see Table 1).

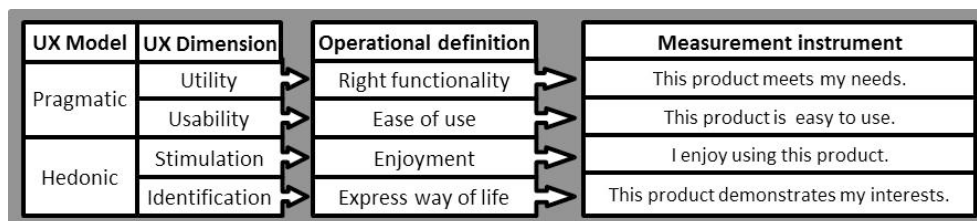


Fig.1: Development of questionnaire items (examples).

The sentence completion section was also developed in two phases. 1) For aiming at good content validity, the groups of sentence stems were created to correspond to UX dimensions of hedonic/pragmatic model (Hassenzahl, 2003). The goal of a sentence stem is to trigger stimulation for participants, which can produce answers in related UX dimension, but also provide unexpected responses. We aimed at easy to understand and relatively unstructured stems encouraging to variety of expression (Holsopple & Miale, 1954). The instructions to fill in the sentence completion tasks were formulated as: "Please, continue the sentences so that they describe you in the best way. There are no wrong answers. Respond rather quickly without thinking too long. You can leave single sentences without answers if you feel that they do not suit to your situation". The first set of sentence completion tasks consisted of 15 sentences. 2) It was refined after a review and filling in by three experts and a naïve representative of target group.

Finally, pre-testing of the method aimed at verifying internal validity of the tool such as: functionality of online survey tool, duration of the filling survey and clarity of the statements. Two respondents were sent email including instructions and a hyperlink to online survey to be filled. The implemented survey contained altogether six pages: 3 for instructions + demo/psychographics, 2 for statements and one for sentence completion tasks. The duration for filling the survey stayed within a convenient time frame (14min) (Asiu et al., 1998; Handverk et al., 2000). The testing resulted in rephrasing of one sentence and adding one new. At the end, the tool consisted of 24 statements and 12 sentence completion tasks.

Case Study

The goal of the study was to explore UX of novel premium sport watch users. The product was aimed at men who wanted to combine style and function: it was designed to be suitable both in business and sports use with features to support outdoor activities. The convenient sampling method was used in recruitment. The

email invitations were sent for registered product users (via American and Italian websites of the company) who agreed to be contacted in research purposes. The invitation and survey were sent either in English or Italian. The Italian answers were later translated into English. As a gift for participation, the respondents took part in a product lottery from the same company. The response time was three weeks and a reminder was sent at the beginning of the third week.

35 participants, men responded to the study (77% Italy, 23% USA). The largest age group was 35-44 years old (42 %). The majority of the respondents had owned the product for 1-6 months (70%), used it regularly (79%), both in hobbies and as a daily watch (58%), and owned another product from same company (49%). The response rate 56% (American sample 47%, Italian sample 59%) can be considered good (Lozar Manfreda et al., 2002). Two responses were removed from the analysis due to incomplete data on all sentence completion tasks.

Results

Statements -Mean and standard deviation for each item and UX dimension are presented in Table 1. Internal consistency of each UX dimensions (Cronbach's $\alpha > .8$), shows good reliability for this study (DeVellis, 2003). The results show the product is perceived positively in 92% of all items, and high score for Overall judgment (mean >4). According to mean scores, Usability and Stimulation were the most successful dimensions, followed by Utility and Stimulation. The product is perceived aesthetically and ergonomically pleasing and of high quality with good usability. It enables to fulfill (most of) the user's needs and provides pleasure and enjoyment. However, the product does not succeed as well to portray personality or to support identification and affinity with other users of the product.

Sentence completion tasks - The analysis was conducted using an adapted version of data-driven content analysis in five phases to identify the central aspects of UX per sentence stem (Krippendorff, 2004): 1) Open (in-vivo) coding was used to create categorization for data of each sentence stem, 2) the frequency per category were counted, 3) categories were labeled as positive, negative or neutral according to its content (no other higher level categories were formed to prevent losing detailed information about UX), 4) All phases of analysis was conducted by one researcher and reviewed by another researcher to improve reliability, finally 5) Analysis of missing data as indicator of internal validity was conducted to identify the least filled sentence stems and effect of fatigue (see N/A rates in Table 1).

A total of 305 sentences were completed in this study (mean response rate 77%). Valence - The positively phrased sentence stems produced mainly positive endings describing factors of positive UX, and negatively phrased stems mainly negative endings describing factors of negative UX, even though there were cases of contrary or neutral endings (see Table 1, e.g. sentences no. 8 and 10). Thematic categorization - Each sentence completion task produced on average six thematically different answer categories (range 3, 9). E.g. the themes of "I am happy with this

product, because..." ranged from brand, status and remembrance issues to appearance, quality and need fulfillment (Figure 2). The most completed sentences characterized the appearance (no. 2), instructions of use (no. 5) and aspects of the product that are considered best and arouse positive emotions (no. 7). The amount of N/A answers does not show a tendency to increase as a function of presentation order of the sentence stems (see the N/A rates in Figure 3).

Tab.1: Mean values with standard deviations for questionnaire items (Likert-scale 1 strongly disagree to 6 strongly agree) and convergence/divergence analysis of the results produced by the methods.

UX Dimension				Utility	Usability			Stimulation			Identification				
				Sentence number											
				2	3	11	1	4	8	5	6	7	12	9	10
PRAGMATIC mean 4.7 (1.33) Cronbach's α .914	Utility mean 4.4 (1.41), Cronbach's α .911	Statement	Mean (std dev)												
		2. This product meets my needs.	4.8 (1.03)	x		x			x			x			
		3. I can rely on this products performance.	4.7 (1.40)												
		4. This product helps me to remember my experiences.	4.2 (1.43)				x					x			
		5. This product enables me to follow up with my performance in my activities.	3.7 (1.64)												
	Usability mean 4.9 (1.14) Cronbach's α .766	6. This product meets my expectations.	4.6 (1.27)						x						
		1. This product was easy to take in use.	4.9 (1.17)					x							
		6. This product provides me information in an understandable way.	4.9 (1.06)				x					x	x		
		7. This product is easy to use.	4.8 (1.21)				x		x						
		HEDONIC mean 4.6 (1.30) Cronbach's α .940	Stimulation mean 4.9 (1.19), Cronbach's α .876	9. This product looks good.	5.5 (0.62)							x		x	
10. This product fits comfortably on the wrist.	5.4 (0.74)							x							
11. It is a pleasure using this product.	5.2 (1.06)						x				x		x		
12. Using this product makes me feel confident.	4.5 (1.37)														
13. I find this product inspiring.	4.3 (1.27)												x		
Identification mean 4.3 (1.35) Cronbach's α .916	14. This product encourages me to challenge myself.		4.1 (1.27)												
	15. I enjoy using this product.		5.2 (1.04)				x				x		x		
	16. This product shows who I am.		4.1 (1.27)												
	17. This product demonstrates my interests.		4.4 (1.27)												
	18. I feel a bond with people who also use this product.		3.9 (1.32)												
Overall judgment mean 4.9 (1.26) Cronbach's α .935	19. This product stands for my way of life.	4.3 (1.21)	x								x				
	20. I feel attached to this product.	4.3 (1.51)													
	21. I like to show this product to other people.	4.7 (1.35)													
	22. This product is worth the money paid.	4.5 (1.23)													
	23. Overall, this is a high quality product.	5.2 (1.13)				x					x				
	24. I could recommend this product to a friend or colleague.	5.0 (1.36)													
TOTAL (Cronbach's α .965)			4.6												
Number of thematic categories per sentence				5	4	7	6	3	6	6	6	9	6	8	7
Percentage of similar themes between the statements and categories of sentence completion tasks (%)				40	0	29	83	33	67	17	67	56	50	0	0
Percentage of original thematic categories of sentence completion results (%)				60	100	71	17	67	33	83	33	44	50	100	100

Comparison of the methods - Convergence of content of the two methods was conducted by cross-tabulation of themes of sentence completion task against statement items (Table 1). The results showed both divergence and convergence between the methods; the sentence completion 1) covered partly the content of statements, (see % of similar themes per sentence in Table 1), and 2) extended the themes of statements (see % of original themes per sentence in Table 1). The sentences aiming to probe Stimulation dimension of UX (sentences no. 5, 6, 7 and 12) produced most convergent results, whereas the sentences probing about Identification (sentences no. 9 and 10) and a negatively phrased sentence probing about Util-

ity (no. 3), were the most divergent. Overall, the sentence completion results divergent from the statements fall into three main thematic categories: 1) detailed descriptions of product qualities (either negative, e.g. it is heavy and the face has poor readability in certain light, or positive, e.g. it is multifunctional, elegant, and provides safety), 2) detailed descriptions of use context (either suitable, e.g. in outdoor activities, or unsuitable, e.g. diving, boxing or other rough context), and 3) descriptions of user and social context, status and the user (e.g. the product attracts attention from others, and gives status to its user).

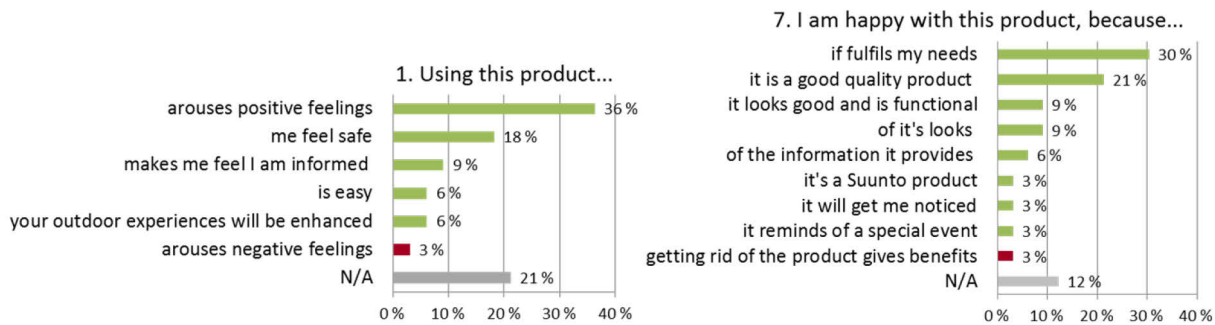


Fig.2: Answer categories of sentences no. 7 and no. 1

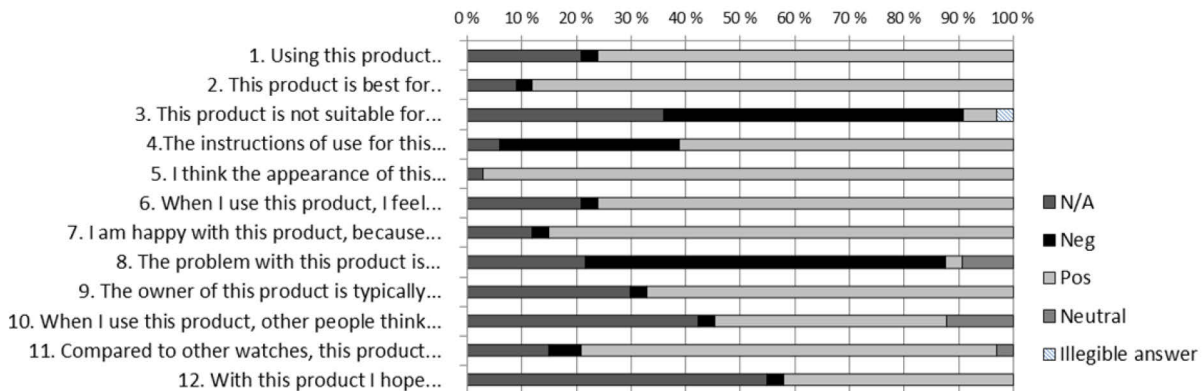


Fig.3: Distribution of the sentence endings to N/A, negative, positive, neutral and illegible answer categories.

Discussion and conclusions

The goal of this paper was to develop mixed method for assessment of remote online UX and to evaluate it. The iteratively developed method was composed of quantitative statements and qualitative projective sentence completion task. In our case study, conducted with premium sport watch users from two language samples, the method was used and meta-analysis was conducted to evaluate it.

The results showed convergence, compensation and complementation between quantitative and qualitative methods providing comprehensive approach to UX evaluation. Convergence was identified as similar valence and themes in both results describing the credibility of the results and content validity of the method. The methods compensated each other as the results of sentence completion provided reasons and explanations for the quantitative UX score. The negatively phrased

sentences especially provided information to guide product development by pointing out causes of dissatisfaction. The methods also provided completeness of phenomenon by covering the universal hedonic/pragmatic dimensions of UX (Hassenzahl, 2003), and providing detailed user-, product- and context of use specific information relevant for the user. For example, in the results of sentence completion, safety came up as an important UX factor in this product context.

There are five suggestion for further development of the method to improve its validity, reliability, and applicability for studying UX. 1) The method needs to be tested for evaluation of broad quality range of the product (e.g. low - high quality, novel - well-established, and personal - organization products, and heterogeneous profile in UX dimensions). This study was limited to the high quality personal sport product. 2) The order effect between two sections of the questionnaire needs to be further investigated. In this study, the statements were completed prior to the sentence completion. Exploration of the opposite or partly integrated order of the sections could reduce the possible effect of respondents' maturation. 4) The procedures and practices for combining and reporting the results of mixed methods need to be studied. The focus could be in identifying the meaningful depth of the results (e.g. strong tendencies averaged across the participants, or individual level weak signals) keeping in mind the simplicity for a lightweight remote UX tool. 5) The method needs to be systematically tested with other comparable methods (e.g. questionnaires with open-ended questions). The extensive comparison criteria including economy, excellence, implementation and assessment related dimensions would provide practitioners detailed information about the characteristic, benefits and limitations of the method and support a long-term safe development of the methods in the field (Kunze et al., 2011).

Acknowledgments

We thank Tiina Taskinen, Christian Nordström from Suunto for co-operation and enabling of the study and professor Timo Saari for his comments on an earlier manuscript. This research was supported by the Finnish Doctoral Program in User-Centered Information Technology (UCIT) and the Finnish Funding Agency for Technology and Innovation (Tekes) through the project SUXES.

Literature

- Asiu, B. W., Antons, C. M., & Fultz, M. L. (1998, May). Undergraduate perceptions of survey participation: Improving response rates and validity. *Paper presented at the meeting of The Annual Meeting of the Association of Institutional Research*, Minneapolis.
- Bargas-Avila, J., A. & Hornbæk, K. (2011). Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proc. CHI'11*. ACM Press, 2689-2698.

- Barkhuus, L. & Rode, J. (2007). From Mice to Men-24 years of Evaluation in CHI. In *Proc. CHI '07* ACM, New York, NY, USA. doi=10.1145/1240624.2180963
- Bruun, A., Gull, P., Hofmeister, L., and Stage, J. (2009). Let your users do the testing: a comparison of three remote asynchronous usability testing methods. In *Proc. CHI'9*. ACM Press, 1619-1628.
- DeVellis, R.F. (2003). *Scale Development – Theory and Applications*. 2nd edition. Sage Publications.
- Evans, J.R. & Mathur, A. (2005). The value of online surveys. *Internet Research* 15 (2), 195-219.
- Hagen, P., Robertson, T., Kan, M., & Sadler, K. (2005). Emerging research methods for understanding mobile technology use. *Proc. OZCHI'05*. ACM Press , 1-10.
- Handwerk, P., Carson, C., & Blackwell, K. (2000). On-line versus paper-and-pencil surveying of students: A case study. *Paper presented at The Association of Institutional Research Conference*, Cincinnati, Ohio, US.
- Hartson, H. R., Castillo, J. C., Kelso, J., Karnler, J., & Neale, W. C. (1996). Remote Evaluation: The Network as an Extension of the Usability Laboratory. In *Proc. CHI 1996*. ACM Press, 228-235.
- Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction*, 19(4), 319-349.
- Hassenzahl, M. (2003). The thing and I: Understanding the Relationship Between User and Product. In: M.A. Blythe, A.F. Monk, K. Overbeeke and P.C. Wright (eds.), *Funology: From Usability to Enjoyment*. Dordrecht: Kluwer. 31–42.
- Hassenzahl, M. & Tractinsky, N. (2006). User Experience - a research agenda. *Behavior & Information Technology* 25(2), 91-97.
- Holmsope, J. Q. & Miale, F. R. (1954). Sentence completion: A projective method for the study of personality. Springfield, IL.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26.
- Krippendorff, K. (2004). *Content Analysis: An Introduction to Its Methodology* (2nd ed.). Thousand Oaks, CA: Sage.
- Kunze, K., Strohmeier, D., & Jumisko-Pyykko, S. (2011). Comparison of two mixed methods approaches for multimodal quality evaluations: Open Profiling of Quality and Conventional Profiling. In *Quality of Multimedia Experience (QoMEX)*, 137- 142. doi: 10.1109/QoMEX.2011.6065691
- Lang, P.J. (1980). Behavioral treatment and bio-behavioral assessment: computer applications. In: Sidowski, J.B., Johnson, J.H., Williams, T.A. (Eds.), *Technology in Mental Health Care Delivery Systems*. Albex, Norwood, NJ, 119–139.

- Lavie, T. & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies*, 60(3), 269–298.
- Law, E. & van Schaik, P. (2010). Modelling user experience - an agenda for research and practice. *Interacting with Computers*, 22(5), 313-322.
- Lozar Manfreda, K., Batagelj, Z., & Vehovar, V. (2002). Design of Web Survey Questionnaires: Three Basic Experiments. *Journal of Computer-Mediated Communication* 7(0).
- Nurkka, P., Kujala, S. & Kemppainen, K. (2009). Capturing Users' Perceptions of Valuable Experience and Meaning. *Journal of Engineering Design*, 20(5), 449-465.
- Riegelsberger, J. & Nakhimovsky, Y. (2008). Seeing the Bigger Picture: A Multi-Method Field Trial of Google Maps for Mobile. In *Proc. CHI'08*, ACM Press, 2221-2228.
- Root, R & Draper, S. (1983). Questionnaires as a Software Evaluation Tool. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, 83-87.
- Schifferstein, H., & Zwartkruis-Pelgrim, E. (2008). Consumer-Product Attachment: Measurement and Design Implications. *International Journal of Design 2* (3).
- Soley, L. & Smith, A. (2008). Projective techniques for social science and business research., The Southshore Press, Milwaukee, USA.
- Strizhakova, Y., Coulter, R.A., & Price, L.L. (2008). The meanings of branded products: A cross-national scale development and meaning assessment. *International Journal of Research in Marketing* 25(2), 82-93.
- Tashakkori, A., & Teddlie, C. (2008). Quality of inferences in mixed methods research: Calling for an integrative framework. In M. M. Bergman (Ed.), *Advances in mixed methods research*. London: Sage.
- Tsai, S-P. (2005). Utility, cultural symbolism and emotion: A comprehensive model of brand purchase value. *Int. Journal of Research in Marketing* 22(3), 277-291.
- Vermeeren, A.P.O.S., Law, E.L-C., Roto, V., Obrist, M., Hoonhout, J. & Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: current state and development needs. In *Proc. NordiCHI 2010*. ACM Press, 521-530.
- Väänänen-Vainio-Mattila, K., Roto, V. and Hassenzahl, M. (2008). Now Let's Do It in Practice: User Experience Evaluation Methods in Product Development. In *CHI EA '08*. ACM, New York, NY, USA, 3961-3964. Doi:10.1145/1358628.1358967

A mixed methods approach to a longitudinal study of user experience in the metals and engineering industry

Heli Väättäjä¹ and Tuomas Oja¹

Keywords: user experience, longitudinal study, manufacturing management system, mixed methods

Abstract

This paper describes a mixed methods approach and its practical application in a longitudinal study of user experience of a manufacturing management system. A mixed methods approach was used in data collection at three phases: 1) initial impressions when the system was deployed at the factory, 2) impressions and experiences in the learning phase, and 3) impressions and experiences in sustained use. Mixed methods approach provided understanding of the factors that can contribute to user experience at different phases. The qualitative and quantitative strands were used to complement, corroborate and expand the findings from the other strand. The used combination of methods was light weight and easy to use, and the results could be easily communicated to different stakeholders. The bidimensional evaluation method is a promising practical tool to capture changes in user satisfaction of different qualities as well as on the levels of the studied descriptive attributes.

Introduction

Suppliers in the metals and engineering industry are turning to usability and user experience as a new source of competitive advantage (Väättäjä et al. 2013). When designing and developing systems that take into account the experiential aspects, it is important to carry out studies on user experience in the natural context of use. Evaluation with the users in long-term use can be used to verify the success of the user-centered design and development activities (ISO 9241-210:2010) and that the produced system quality satisfies the users and meets the identified needs and set requirements.

Complex systems are often customized in every delivery. This is also the case for manufacturing management systems, which control concurrent units consisting of robots, machine tools, and computers. Prior to installation and system deployment at the customer's facilities, it is in practice impossible to test and evaluate a complex system as a working entity. This is due to the physical size of the complex system as well as due to the fact that the components of the complex system entity are delivered by several subcontractors. Iterative testing of non-functional and functional prototypes starting from the early phases of the user-centered design

¹ Tampere University of Technology, Department of Pervasive Computing, Tampere, Finland

process is needed to ensure user feedback to guide design and development activities. Furthermore, user experience studies in real life settings provide information on factors that can contribute to user experience that cannot be captured otherwise.

In this study we aimed to capture user experience and identify contributing factors within the customer's manufacturing context after the delivery of the system to the customer. Our study aimed to identify, develop, and trial a toolbox of light-weight and easy to use methods, that could later be applied in field conditions by the supplier, and that could with relatively little effort provide the supplier information they could use in their R&D activities. Furthermore, the methods should also be useful for research purposes as the aim was user experience in a longitudinal study. This paper presents the mixed methods used in the data collection, practical issues to consider when carrying out studies in this domain, as well as proposes how the used methods can be used in practice in field studies and evaluation.

Mixed methods approach in data collection

To support the evaluation and to identify the factors contributing to user experience, we chose to use a mixed methods approach (Tashakkori & Teddlie, 2008) in the study. Qualitative and quantitative strands were used to complement, corroborate and expand the findings from the other strand of the study (ibid.). They provided a more comprehensive and more complete account of the area of the study and understanding of the context, enabled finding explanations for the findings, as well as augmented the findings of one strand by the other strand (Bryman, 2008). The suitable combination of methods for each phase was chosen to support creating an understanding of user experience and capturing variations in evaluative judgments as well as to support development activities, especially in case an issue was identified that needed attention. Contextual interviews and observations were carried out with a varying focus depending on the phase of the study, and a paper-based questionnaire was administered at all phases.

Qualitative data was collected with the following methods:

- *Contextual interviews* on the shopfloor level were carried out to capture users' open descriptions of impressions on the system and experiences.
- *System usage was observed* to identify factors contributing to user experience at all phases. To capture and communicate to developers a breakdown in the task flow in the second phase, *the work was modelled* based on the observation data *by a sequence model* to capture the real ordered steps that the user takes when performing the tasks (Holzblatt et al. 2005).
- *Observations were supported by recording a short video of the usage* with the think-aloud method while user was performing his task with the system. This helped in creating the sequence model of the workflow, specifically in the case, where a breakdown was identified. Video also helped in communicating the breakdown easily to the developers.

- *A sequence of steps* by the user when carrying out a task *was visualized on a floor map* in the second phase of the data collection (Holzblatt et al. 2005). This visualization provides information of the actions and tasks of the user and their patterns in relation to the context of use with machinery and other artifacts and how artifacts, technology, information and other users are situated in the physical context.
- *A paper-based questionnaire* was administered at each phase of the data collection. It applied the themes of the interview questions by asking users to fill in sentence completions in the questionnaire (Soley et al., 2008) as well as with open field questions: 1) "Describe freely your impressions of the system in 1-3 ways: The system is in my opinion..." (three open lines to fill), 2) "What aspects did you pay attention to in the system? I paid attention to..." (three open lines to fill), 3) "What other thoughts, feelings, or ideas does the system raise in you?" (three open lines to fill). In addition participants were asked to describe freely from one to three most important strengths as well as weaknesses of the system, and for their impressions on impacts of the system on their work.
- *Quantitative* data was collected with the following methods:
- We used in the paper-based questionnaire the *bidimensional evaluation method* (Jumisko-Pyykkö et al. 2008) combining satisfaction and acceptance measurements for evaluation of the experiential qualities (attributes). Satisfaction was measured on an 11-point scale (0-10) combined with a binary yes/no assessment for the acceptability of the quality (ibid.). This measurement provides a clear indication of the satisfaction with and acceptability of system qualities and the change at different points of measurement.
- We included *items from UTAUT* (Venkatesh et al. 2003) into the questionnaire, to assess the enjoyment to use the system, and having the necessary knowledge to use the system using a seven-point Likert scale. In addition, we included an item about recommendation to colleagues, with the same scale.
- In the second and third phase we used the *Systems Usability Scale* (SUS, Brooke 1996) with ten items to provide a basic usability scale measure to compare it to reported benchmark values by Sauro (2011). By comparing the sum of the items multiplied by two (scales to 0-100) to a benchmark gives an indication for the users' judgment on usability. An alternative shorter measure to be considered for field studies is the ASQ (After-scenario questionnaire, Lewis 1991), which can be applied in field studies by using items on satisfaction with the ease of use and the amount of time to complete a task.

Timing of the data collection

The data collection of the longitudinal study was carried out in three phases of the study. Initially, we planned to collect the data at the following time points: prior to deployment (expectations), at the time of deployment right after the training given

to the employees on system and its usage at the customer company (first impressions), after two or three week usage (learning phase), after three months usage (sustained use) and at six months (repeated sustained use). This plan turned out to be unrealistic due to the reasons discussed next.

When planning the research design we found that we were not able to gain access to the employees prior to deployment of the system at the customer company. The first possibility was therefore at the time of deployment when the training was given to the users. This enabled the collection of first impressions.

When discussing with the representative of the supplier's customer company, it turned out that collecting data after a few weeks of deployment to capture the learning phase impressions did not seem optimal. In case of complex systems, the initialization phase can take up to a half a year, as part of the machinery may be installed later and possible issues needing changes and improvements in the implementation. The employees gradually start trialing the system and the related new processes in real life use and taking them into use. This whole period is in fact "learning phase" for the users, until the system is used in daily use. Due to this, and the fact that the contact person was confident that within the first weeks there is very little use of the system in practice, the second phase of the study was carried out three months after the deployment to capture the impressions and experiences.

When carrying out the second phase, we decided to postpone the third phase until about nine months to a year from the deployment. This was due to the fact, that the learning phase was clearly still ongoing. This third measurement aims to capture the impressions in sustained use.

Experiences with the used data collection methods

We next discuss our experiences with the used data collection methods.

Contextual inquiry combined with video recording and sequence modelling proved to be an effective and easy tool in communicating the usability related issues to the supplier company. However, video recording needed special permission from the customer company a priori to entering the field. The supplier's customer company had strict rules that needed to be obeyed regarding recording of photos and videos. This aims to protect their interests in terms of confidentiality of their manufacturing systems and especially their processes. Permission to record was given when researchers agreed to only capture video of the user interface of the supplier's manufacturing system as well as getting approval to take the video clip out of the customer company's premises after it was first checked by a representative of the customer.

Interviews and observations of operators and programmers carried out on the shopfloor level provided understanding of the context of use and experiences of the users. First hand access to users and seeing them use the system was the only way to truly understand what factors can contribute to user experience. Writing field

notes was done to a small paper notebook either directly or when a convenient moment occurred. The field notes were analyzed with a data-driven content analysis, by using open initial coding, grouping the codes and fixing the coding and creating descriptions for the groups. This field note data enabled a comparison to the questionnaire responses as well.

Paper-based questionnaires that were administered during the field visits complemented the interviews and observations. The analyzed results from both open and closed ended questions could be easily communicated to the practitioners at the supplier company. The results from the bidimensional evaluation method with satisfaction and acceptance assessment was a concrete tool for demonstrating the temporal changes both in satisfaction and acceptance of the experiential qualities. In addition, comparing the result of SUS measurement with a benchmark value, concretizes the usability related quality of the system as an overall evaluative judgment. These assessment methods provide support for user-centered design activities and focusing of the iterative development efforts to unacceptable qualities.

Future work could address the application of the bidimensional evaluation method in practical verification of set targets by the company in longitudinal measurements of user experience. In practical work a priori set target levels of assessed quality and target percentages for acceptance and their combinations could be used to verify the success of the user-centered design process. The results could further be used as hard facts in sales and marketing efforts by suppliers.

Discussion and Conclusions

To conclude, the used mixed methods approach was found useful both for the researchers as well as for supporting the development activities in the supplier company. Approach provided insight into the users' impressions, experiences, factors contributing to user experience as well as satisfaction with and acceptance of the experiential qualities. Results were easily communicated to different stakeholders to provide support for user-centered design activities and to focus supplier's efforts to unacceptable qualities. Carrying out user experience related studies in industrial context is challenging, as the assumptions of the researchers, the reality of the industrial production context and practical arrangements need negotiation and special attention in the planning of the studies. Careful planning and flexibility when planning the research design provide a good basis for implementing the studies in practice. Future work could address a larger scale study applying the bidimensional evaluation method in industrial context for user experience studies.

Acknowledgements

This research was supported by TEKES as part of the User Experience and Usability in Complex Systems (UXUS) program of FIMECC and by the Finnish Doctoral Program in User-Centered Information Technology (UCIT). The authors thank the

supplier company as well as the customer company for the possibility to carry out this research.

Literature

- Brooke, J. (1996). SUS: a "quick and dirty" usability scale. In P.W.Jordan, B. Thomas, B.A. Weerdmeester, & A. L. McClelland. *Usability Evaluation in Industry*. London: Taylor and Francis.
- Bryman, A. (2008). Why do Researchers Integrate/Combine/Mesh/Blend/Mix/Merge/Fuse Quantitative and Qualitative Research? In M.M. Bergman. *Advances in Mixed Methods Research*, 87--100. Great Britain: Sage.
- Holzblatt, K., Wendell, J.B., & Wood, S. (2005). Rapid contextual design: A How-to Guide to key techniques for user-centered design. Morgan Kaufmann Publication.
- ISO 9241-210:2010 (2010). Ergonomics of human-system interaction. Part 210: Human-centred design for interactive systems. International Standardization Organization (ISO).
- Jumisko-Pyykkö, S., Malamal Vadakital, K.V., Hannuksela, M.M. (2008). Acceptance threshold: Bidimensional research method for user-oriented quality evaluation studies. *International Journal of Digital Multimedia Broadcasting*, vol. 2008, Article ID 712380, 20 pages. doi:10.1155/2008/712380
- Lewis, J.R. (1991). Psychometric evaluation of an after-scenario questionnaire for computer usability studies: The ASQ. *SIGCHI Bulletin*, 23(1), 78-81.
- Sauro, J. (2011). Measuring usability with the system usability scale (SUS). Available at: www.measuringusability.com/sus.php
- Soley, L. C., & Smith, A.L. (2008). *Projective methods for social science and business research*. The Southshore Press.
- Tashakkori, A.T. & Teddlie, C. (2008) Quality of Inferences in Mixed Methods Research: Calling for an Integrative Framework. In M. Bergman. *Advances in Mixed Methods Research*, 101-119. Great Britain: SAGE Publications.
- Venkatesh, V., Morris, M., Davis, G., & Davis, F. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425-478.
- Vääätäjä, H., Seppänen, M., & Paananen, A. (2013). Creating value through user experience: A case study in the metals and engineering industry. Accepted to be published in the *International Journal of Technology Marketing*.
- Wixom, B. H., & Todd, P. A. (2005). A theoretical integration of user satisfaction and technology acceptance. *Information systems research*, 16(1), 85-102.

A Collaborative Approach for the Preparation of Co-operative Multi-User Driving Scenarios

Julian Schindler¹, Johann Kelsch¹, Matthias Heesen², Marc Dziennus¹, Gerald Temme¹ and Martin Baumann¹

Keywords: exploratory scenario design, collaborative platform, multi-touch table

Abstract

The design of scenarios for driving simulators gets more and more complex due to the upcoming need of creating scenarios where different types of interactions are tested. This includes the driver-vehicle as well as driver-traffic, driver-driver or vehicle-vehicle/vehicle-infrastructure interactions, e.g. by using V2X technology. This paper describes a new collaborative approach as part of a PhD-thesis making it possible to design such scenarios in short time: The Exploratory Scenario Design. The theoretical background of the approach is sketched as well as the current step of implementation, using e.g. a multi-touch table.

State of the Art Driving Scenarios

Since the first development of computer systems such systems were used to simulate complex system behaviors as simulation has a very high reproducibility and standardization, making it possible to alternate parameters and to isolate specific effects, often in a cheap way without any hazards.

During the last decades, and esp. since computer graphics were able to produce virtual realities with a high level of realism, this became also true for the automotive domain. In order to understand drivers and their abilities, driving situations and their challenges, or driving phenomena at all, it was necessary to perform lots of driving simulator experiments with a big number of variations.

In the first years of driving experiments in virtual worlds, driving scenarios were simple: As the simulators lacked realism the first simulated driving scenarios using computer generated graphics according to Pollock et al. (1999) were highway or rural road scenarios without any traffic (e.g. Witt & Hoyos (1976) or Donges (1978)). As computer performance rose, it became possible to realize other vehicles and afterwards more complex traffic behavior and traffic situations in more complex environments, leading to first scenarios in urban conditions.

In the same time, driver assistance systems and afterwards Advanced Driver Assistance Systems came onto the market, obtaining more and more control over the vehicle and interacting more and more with the drivers. From now on, driving scenarios had to cope not only with the interaction between driver and car or driver

¹ Deutsches Zentrum für Luft- und Raumfahrt, Institut für Verkehrssystemtechnik

² Fraunhofer-Institut für Kommunikation, Informationsverarbeitung und Ergonomie FKIE

and traffic, but also with the interaction between the driver and the assistance system on board.

Current scenarios go even beyond: As modern ADASs not only focus on the behavior of the host car, but aim at the interaction between road users or infrastructure, directly or via V2X, the focus of driving scenarios changed accordingly.

Furthermore, it has been found that people behave different in pure simulation environments than in environments with both human and computer-controlled vehicles (Winner & Hermann, 2012).

DLR's driving simulation environment kept up with the times: Starting with simple fixed-based simulators, the realism was raised up to a motion-based simulator. Current scenarios with the focus on interaction between road users can be implemented using the new Modular and Scalable Application Platform for ITS Components (MoSAIC, see Lorenz et al., 2011). In MoSAIC, the different driving simulators are connected to each other, making it possible to test a wide spectrum of interactive scenarios with multiple human drivers.

The MoSAIC setup is currently in use at various international projects, e.g. the European Artemis SP8 D3CoS project (see Heesen et al. 2012).

Challenges in Scenario Design

At the beginning of any driving scenario design, there is an interesting effect that stimulates a research question, and a hypothesis about the causal mechanism for this effect as a possible answer for this research question. This is combined with the expectation that this research question can be answered by performing experiments with subjects. The general challenge of scenario design, as in any experimental design, is the transformation of the research question into a sound experiment. As shown in figure 1, in a first step the research question has therefore to be transformed into a rough idea of a possible driving situation.

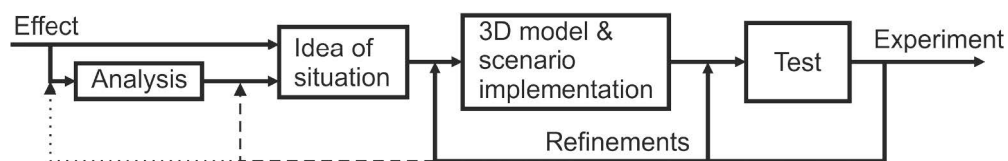


Fig. 1: Common scenario design

Whenever there is a first idea of a situation, this situation has to be modeled in detail, including the environment of the situation, i.e. the road network, buildings etc., and the behavior and point of (dis-)appearance of the road users surrounding the subject. Esp. this part of the transformation is challenging due to the following reasons:

- *Level of detail:* A modeled situation needs the specification of exact road user positions, velocities, etc., mostly unknown to the requester of a scenario.

- *Generalization*: The requested situation should occur in any case, fully independent of the driving behavior of the subject.
- *Involvement of different people*: The requester of a scenario shown in a driving simulator is commonly not the implementer of the model. Therefore a requirement specification has to be formulated which mostly (according to the first point) lacks details. In addition, the requester and the implementer often have different backgrounds, which may lead to misunderstandings or different interpretations of the specification.
- *Time*: A realistic scenario needs often more time for preparation as estimated, esp. due to the involvement of different people or the lack of knowledge on features of the used software.
- *Virtualization*: Driving simulators differ in terms of the grade of reality between each other and compared to real driving situations. Therefore, results of other studies may not be adopted wholesale and will need time consuming fine tuning.

When the scenario is implemented, it needs to be tested best by one or more persons other than the implementer. During these tests, in most cases issues are found in the scenario and it needs to be fixed or even re-implemented with the mentioned challenges, making the whole procedure very time-consuming.

While this is already true for simple scenarios it becomes a real challenge when handling current complex scenarios like those sketched in the beginning, using e.g. cooperation between the subject vehicle and its surroundings or even more than one subject vehicle. Esp. those multi-user scenarios create many new challenges, as the reproducibility of the scenario stays in conflict with the individual behavior of the different drivers, esp. when they have to meet each other in one special situation after a while of driving.

Existing Approaches

Some of the above mentioned points have already been addressed by the major suppliers of driving simulator software: Instead of bothering the implementers with low level details of vehicle behavior when writing the scenario scripts, often a higher level language is used including virtual driver skills and vehicle skills or maneuvers, applying an artificial intelligence to the virtual vehicles (see e.g. Olstam (2009) or VIREs (2006)). Instead of using scripts at all, which can be hard to read and which require a special and mostly non-standard syntax to be learned by the implementers in advance, most driving simulator software is already shipped with a scenario editor providing a Graphical User Interface (GUI).

These two points already lead to a tremendous speed-up in the scenario design, as e.g. Drag-and-drop is supported by the GUIs, the learning phase is shortened, the skill requirements are reduced and therefore the software can be used by more people. On the other hand, the creation of artificial intelligence in vehicle behavior

may result in new challenges as those vehicles tend to behave always correctly esp. in terms of traffic regulations, which often differs from the requirements of the scenario, as non-conformant behavior is the key to many effects of driving.

In order to avoid the problems with fully autonomously controlled vehicles it is of course possible to manually control all the road users surrounding the subject vehicle, like e.g. already done by Currie (1969). This brings a huge benefit in flexibility but lacks in terms of reproducibility. Therefore, Olstam & Espié (2010) suggested to control the traffic cars manually in predetermined situations at which measurements are taken and to shift the control in a non-conspicuous way to a fully autonomous mode in-between. This “theatre metaphor” leads to a better reproducibility while still keeping the flexibility, but additional trained personal and infrastructure is needed to drive the vehicles.

Independent of using manually or autonomously controlled vehicles, as the requester of a scenario often does not equal the implementer as described above, the problem of misunderstandings and different interpretations of possibly too vaguely formulated requirements for the scenario design can only be reduced when tools promote the mutual understanding of the desired scenario, for example by facilitating the discussion between both parties. Tönnis & Klinker (2009) developed a collaborative approach for discussion and development of traffic scenarios by using a table-top platform. In their approach, real toy cars could be moved on a table-top display showing a bird view of the scenario, directly creating a virtual representation moving in the same way in an attached driving simulator in driver’s view. Trajectories from start to end could be recorded and replayed, making it possible to iteratively build up a set of fixed trajectories for all autonomously controlled vehicles, resulting in a complete scenario. See also Tönnis (2007) for further details.

Although this kind of setup is very intuitive, its approach is limited in the fact that subjects do not behave equally in every situation, making pre-recorded trajectories unusable when the delta between the assumed position of the subject and the real position gets too high. The possibility of the platform to directly sketch the scenario by manually controlling vehicles on the table while a subject is driving has its pros and cons as described before.

When thinking of scenarios with several human drivers this approach comes to its limitations very fast, as the display size is limited and therefore it is impossible for the “implementer” to control cars spread in the virtual world, and the number of surrounding vehicles will increase. The key challenge is to develop a collaborative approach which copes with the diversity of subjects, focusses on a good usability while still having the ability of changing every low level detail which might be of interest in the scenario design.

A new Collaborative Approach

In order to deal with this challenge, the common scenario design shown in figure 1 has to be adapted. This is esp. true for the implementation and testing phases of the design, as they have been identified as difficult and very time consuming. A possible solution of this has been found in the Exploratory Design (Flemisch et al., 2008) as used for ADAS development (e.g. Schindler et al., 2010). Within the Exploratory Design, the user of a system gets directly involved into the design process. In terms of scenario design this can be achieved by the collaboration of the requester, implementer and tester of the scenario as early as possible.

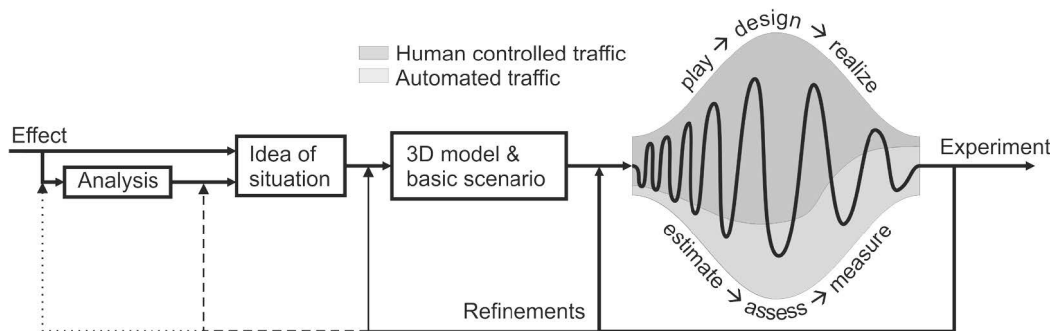


Fig. 2: The Exploratory Scenario Design process adapted from Flemisch et al. (2008)

As shown in figure 2, a basic 3D model where the situation is generally planned to take place is needed in advance, i.e. a model of an urban, rural or highway area, combined with a basic scenario, e.g. consisting of mean traffic densities only. Starting with this setup, the detailed scenario design can start.

In order to promote the collaboration, an Ideum MT 55" Multi-Touch-Table with a maximum of 32 parallel touch points has been equipped with a bird view perspective on the driving scenario. It serves as a fully functional graphical scenario design tool by making it possible to easily insert traffic cars into the scenario by Drag-and-drop gestures and to change their parameters, e.g. their routing information or their behavior, in low level or higher level format at any time by simply pointing at the corresponding menus with a finger. As the Multi-Touch-Table is a table for discussion, it is controllable from every place around the table, including the possibility to pick shown menus or program interactions, and to rotate and translate them on the screen as needed.

Furthermore it is capable of showing the scenario online with the possibility to directly control the vehicles on the table with a fingertip. In difference to the approach of Tönnis, no "tangible cars" (Tönnis, 2007) are used on the table, bringing the benefit that a car can be driven with one single finger instead of controlling it with one hand. The driving is done in a way that one finger drags to the target-point the vehicle aims to go, and the vehicle moves there under its physical restrictions. As the vehicles are virtually present only, there are no scaling restrictions on the table, making the zoom into the scenario continuously adjustable.

Alternatively, as the implementation is part of the DOMINION software framework (Köster et al., 2008), any traffic vehicle can be controlled by mouse, keyboard, any standard joystick or game wheel, or by attaching any complete driving simulator. In the current implementation, up to 32 vehicles can be theoretically controlled on the table while at the same time up to four vehicles can be controlled by driving simulators or other control devices. The driving simulators can additionally be used to get an impression of the perspective the subject will have.

Due to the lack of exact reproducibility when controlling many vehicles manually, one of the key features of the Exploratory Scenario Design is the transferability of the control to an autonomous mode. In difference to the Tönnis approach, not only trajectories, but also meta-data of the manually driven vehicles is recorded. This meta-data includes e.g. aspects of the traffic situation, for example distances and time-headways to other road users, and behavioral data in terms of driven maneuvers and its parameters.



Fig. 3: Scenario preparation around the Multi-Touch-Table at the DLR MoSAIC Lab

The recorded data can afterwards be replayed directly, promoting the discussion. Nevertheless, the major benefit is that it facilitates the transfer into automated traffic vehicle behavior on a higher level, as it gives hints on the conditions where a driving maneuver should occur and on the parameters of this maneuver. Therefore, the scenario designers can directly pick the relevant conditional parameters of the situation, change the constraints if necessary, combine them with a maneuver to be performed and specify the tolerated deviation of the maneuver execution parameters. By this it is possible that even automated traffic vehicles can adjust their behavior according to different subject behavior in well-defined borders.

The different steps of the Exploratory Scenario Design process can be iterated as needed, leading to a crisp scenario design already after a minimum of time.

The shown tools are currently under preparation and are already partly usable (see figure 3). The current implementation stage already consists of the touch

framework and the (meta-)data recording, already speeding up the scenario design in a perceptible way. Further implementations will follow.

Conclusion

It has been shown that scenario design is getting more and more complex and therefore challenging. The approach of Exploratory Scenario Design helps to solve this issue by using a multi-touch table which facilitates discussions. In presenting the possibility of attaching driving simulators and different input devices directly, the testing of the scenario is already included in the design phase. While scenario data including meta-data is recorded during the design phase, vehicle behavior can directly be adapted as needed.

References

- Currie, L. (1969). The Perception of Danger in a Simulated Driving Task. In *Ergonomics* 12 (6), pp. 841–849
- Donges, E. (1978). A two-level model of driver steering behavior. In *Human Factors: The Journal of the Human Factors and Ergonomics Society* 20 (6), pp. 691–707.
- Flemisch, F., Schindler, J., Kelsch, J., Schieben, A., Damböck, D. (2008). *Some Bridging Methods towards a Balanced Design of Human-Machine Systems, Applied to Highly Automated Vehicles*. Applied Ergonomics International Conference, Las Vegas (USA)
- Heesen, M., Baumann, M., Kelsch, J., Nause, D., Friedrich, M. (2012). *Investigation of Cooperative Driving Behaviour during Lane Change in a Multi-Driver Simulation Environment*. In: Human Factors: a view from an integrative perspective. Human Factors and Ergonomics Society (HFES) Europe Chapter Conference, 10.-12. Okt. 2012, Toulouse, Frankreich.
- Köster, F., Gacnik, J., Hannibal, M. (2008). *Serviceorientierung als Zugang zur Strukturierung von in-car Softwaresystemen*. In: Informationssysteme für mobile Anwendungen. IMA 2008, Braunschweig
- Lorenz, T., Baumann, M., Jaschke, K., Köster, F. (2011). A Modular and Scalable Application Platform for Testing and Evaluating ITS Components. In: *20th IEEE International Conference on Collaboration Technologies and Infrastructures -CoMetS'11*, ISBN 978-0-7695-4410-6
- Olstam, J. (2009). *Simulation of surrounding vehicles in driving simulators*. Linköping, Univ., Diss., 2009. Linköping: Linköping University (Linköping studies in science and technology : Dissertation, 1248)
- Olstam, J. & Espié, S. (2010). Combination of autonomous and controlled vehicles in driving simulator scenarios. In Andrea Benedetto (Ed.): *Advances in Transportation Studies*. University Roma Tre (21), pp. 23–32

- Pollock, D., Bayarri, S., Vicente, E. (1999). A historical perspective of the use of driving simulators in road safety research, In: *Progress in system and robot analysis and control design*, Springer, London, ISBN 978-1-85233-123-8
- Schindler, J., Temme, G., Schieben, A., Flemisch, F. (2010). *Exploratory Design of a Highly Automated System for Entering the Expressway*. In D. de Waard, A. Axelsson, M. Berglund, B. Peters, C. Weikert (Eds.): *Human factors. A system view of human, technology and organisation*. Maastricht: Shaker Publishing, pp. 201–216.
- Tönnis, M. (2007). The Tangible Car - Rapid Intuitive Traffic Scenario Generation in a Hybrid Table-top and Virtual Environment. *The Fourth International Workshop on the Tangible Space Initiative in conjunction with the Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality*, Nara, Japan, Nov. 13 - 16, 2007
- Tönnis, M. & Klinker, G. (2009). A Collaborative Table-top Platform for Discussion and Development of Traffic Scenarios with Human Behavior. *The 17th World Congress on Ergonomics (International Ergonomics Association, IEA)*, Beijing, China, Aug. 9 - 14, 2009
- VIRES GmbH. (2006). *v-Scenario – Scenario Editor – Product Data Sheet*, available at <http://www.vires.com/Docs/ScEd20061026.pdf>, last accessed Aug. 2013
- Winner, H. (Ed.) (2012). *Handbuch Fahrerassistenzsysteme. Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort*. 2. Aufl. Wiesbaden: Vieweg + Teubner (ATZ-MTZ-Fachbuch).
- Witt, H. & Hoyos, Carl G. (1976). Advance information on the road: A simulator study of the effect of road markings. In *Human Factors: The Journal of the Human Factors and Ergonomics Society* 18 (6), pp. 521–532.

Entwicklung eines Inhome Displays zur Überwachung und Steuerung des Stromverbrauchs

Anke Schwarze, Maximilian Horn, Michael Haiduk und Frank Eggert¹

Schlüsselwörter: HMI, Smart Meter, Stromverbrauch, Display, Verhaltensänderung

Zusammenfassung

Um die Energie- und Klimaziele der Europäischen Union zu erreichen, soll auch der private Stromverbrauch reduziert werden. Außerdem bedarf es einer Verlagerung des Stromverbrauchs in Zeiten mit einem hohen Anteil von Strom aus regenerativen Energiequellen. Somit müssen Haushalte ihr Stromverhaltensverhalten verändern. Verhalten, das in privaten Haushalten mit dem Verbrauch von Strom verbunden ist, ist in vielen Fällen stark routiniert. Eine Veränderung von solchem Verhalten allein durch das individuelle Bemühen stellt sich häufig als sehr schwierig heraus und sollte daher durch externe Reize unterstützt werden. Eine Möglichkeit besteht in der Nutzung eines Inhome Displays, das relevante Anreize zur Verhaltensänderung gibt. Damit ein solches Display Verhaltensänderungen zur Folge hat, müssen die dort präsentierten Inhalte relevant für die menschliche Verhaltenssteuerung sein. Die vorliegende Arbeit beschäftigt sich mit der Identifikation solcher relevanten Stimuli und ihrer Validierung, zunächst im Laborsetting.

Einleitung

Das Ziel der Europäischen Union besteht darin, bis zum Jahre 2020 die Energieeffizienz und den Anteil erneuerbarer Energien um 20 % zu erhöhen und 20 % Kohlenstoffdioxid einzusparen (European Commission, 2010). Studien zeigen, dass der Einsatz von Smart Metern, bzw. die damit möglich werdenden Maßnahmen, den Stromverbrauch reduzieren kann (z. B. Reduktion des peak time-Verbrauchs: Ivanov, Getachew, Fenrick, and Vittetoe (in press)), wobei die Art des Feedbacks entscheidend ist (Darby, 2006). Eine Kombination mit einem Medium zur Darstellung relevanter Verbrauchsinformation und weiteren aufgearbeiteten Daten ist unerlässlich.

Das Projekt EVE – Effiziente VerbraucherEinbindung² beschäftigt sich mit der Frage, wie das Stromverhaltensverhalten von Haushaltskunden mittels eines Inhome Displays beeinflusst werden kann. Es wird unter anderem untersucht, welche Darstellungsprinzipien den Nutzer dabei unterstützen, seinen Stromverbrauch zu reduzieren und/oder zu verlagern.

¹ Technische Universität Braunschweig, Institut für Psychologie, Abteilung Psychologische Methodenlehre und Biopsychologie

² Gefördert vom Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit aufgrund eines Beschlusses des Deutschen Bundestages

Theoretischer Hintergrund

Vor allem im Bereich des umweltfreundlichen Verhaltens zeigt sich häufig, dass Menschen eine positive Einstellung zu umweltfreundlichem Verhalten kommunizieren und Stromspar-Absichten bekunden, sich diese jedoch nicht im tatsächlichen Verhalten zeigen (Huber, 2001). Gründe hierfür können beispielsweise eine mangelnde Information über Umsetzungsmöglichkeiten, höhere monetäre Kosten oder eine Einschränkung des Komforts sein (vgl. z. B. Diekmann & Preisendörfer, 1992; Franzen, 1997). Vor allem bei der Reduktion oder Verlagerung des Stromverbrauchs in privaten Haushalten stehen außerdem tägliche Routinen im Vordergrund, die einer Änderung des Verhaltens im Wege stehen. Da der kurzfristige Benefit von Energieverbrauch im Alltag häufig wichtiger ist als Energieeffizienz, entstehen energieverwendende Routinen mit einer höheren Wahrscheinlichkeit als energiesparende Routinen (Mack, 2007).

Um menschliches Verhalten zu ändern, finden sich in der Literatur verschiedene Möglichkeiten. Im Bereich Umweltschutz wurden und werden zahlreiche unterschiedliche Maßnahmen durchgeführt. Diese umfassen beispielsweise Informationskampagnen (z. B. Staats, Wit, & Midden, 1996), Diskussions-Interventionen (z. B. Werner, Sansone, & Brown, 2008), Feedback-Interventionen (z. B. Staats, Harland, & Wilke, 2004) oder auch Kombinationen verschiedener Maßnahmen (z. B. Laurens, 2012).

Im Kontext des Stromverbrauchsverhaltens bietet sich durch die Technologie des Smart Meters die Möglichkeit, Informationen über den aktuellen Stromverbrauch zeitnah zu übermitteln. Da es Menschen schwer fällt, ihren Stromverbrauch im Auge zu behalten und ihn zu prognostizieren (Gans, Alberini, & Longo, 2013) ist die Darstellung des Stromverbrauchs auf einem Display, das beständig in der Wohnung verfügbar ist, eine erste Unterstützung der Überwachung des Stromverbrauchs. Darüber hinaus besteht die Schwierigkeit darin, dass Elektrizität für viele Menschen sehr abstrakt ist (Hargreaves, Nye, & Burgess, 2010) und keine Verbindung zwischen Verhalten und dem Ausmaß des Stromverbrauchs hergestellt wird. Dies muss bei der Entwicklung von Darstellungsformen berücksichtigt werden. Eine reine Information über den Verbrauch leitet eine Änderung des Verhaltens nicht zuverlässig ein. Das heißt, weitere Hinweise oder Stimuli sollten hinzugefügt werden. In dem Projekt EVE wird ein Inhome Display entwickelt das genutzt wird, um Stimuli darzubieten, die einen Einfluss auf das Verhalten haben können. Für das Verhalten sind unterschiedliche Faktoren relevant, so dass in dem Display ebenfalls unterschiedliche Ebenen adressiert werden.

Darstellungen auf dem Inhome Display

Da Elektrizität an sich und auch die mit ihm verknüpfte Einheit Kilowattstunde (kWh) als sehr abstrakt wahrgenommen werden, sollte der Stromverbrauch zusätzlich auch in den Einheiten Euro und Kohlendioxid (CO₂) angezeigt werden. In einer vorangegangenen experimentellen Laborstudie konnte gezeigt werden, dass für

die Darstellung von CO₂ besonders verhaltensnahe und vertraute Darstellungen präferiert werden (Horn, Haiduk, & Schwarze, 2013).

Um einen Anker für den Stromverbrauch zu haben, kann der Nutzer den aktuellen Verbrauch mit seinem Stromverbrauch zu einem anderen Zeitpunkt vergleichen. So kann der interessierte Kunde beispielsweise Veränderungen durch neue Elektrogeräte oder den Effekt einer Sparmaßnahme selbstständig überprüfen.

Verhalten wird auch durch soziale Vergleichsprozesse beeinflusst. Da dies beim Stromverbrauchsverhalten ebenfalls zu vermuten ist, werden soziale Vergleiche hinsichtlich des Stromverbrauchs ermöglicht. Beispielsweise kann sich der Nutzer mit ähnlichen Haushalten oder Freunden vergleichen.

Um eine längerfristige Nutzung zu fördern, können individuelle „Stromsparziele“ gesetzt werden. Diese können beispielsweise in Relation zu anderen sozialen Gruppen formuliert werden oder im Hinblick auf den eigenen bisherigen Verbrauch. Ein kontinuierliches, einfach wahrnehmbares und schnell zu verarbeitendes Feedback über den Stand der Zielerreichung soll das Verhalten steuern. Die ständig präsente Rückmeldung der Konsequenzen des Verhaltens soll es ermöglichen, alte Verhaltensmuster zu überwinden und neue, an die Ziele angepasste, zu etablieren. Die Rückmeldung erfolgt über etablierte Reize, die eine bestimmte Bedeutung erlangt haben und nicht mehr beliebig zu nutzen sind (z. B. grün für "alles in Ordnung", rot für "nicht alles in Ordnung").

Die Funktionalität des Displays wird weiter ausgebaut, beispielsweise werden besonders hohe Verbräuche farblich hervorgehoben. Eine große Auswahl von Funktionen und Individualisierbarkeit sollen die Akzeptanz der Nutzer erhöhen.

In mehreren Iterationen soll das Display so gestaltet werden, dass sowohl der Effekt auf das Verhalten als auch die Akzeptanz möglichst groß werden. Das Ziel der ersten Laborstudie besteht darin, einen Eindruck der Usability der App zu erhalten und Hinweise für die weitere Entwicklung von den Probanden zu erlangen.

Methode¹

In der Studie haben die Probanden die Aufgabe, eine erste Version des Inhome Displays zu evaluieren. Anhand objektiver und subjektiver Daten wird die Usability untersucht.

Das Inhome Display wurde als App auf einem Tablet Computer umgesetzt. Das Graphical User Interface (GUI) des Displays wurde mit dem Android Development Toolkit und Eclipse programmiert und wird auf einem Lenovo ThinkPad Tablet (NZ72DGE) präsentiert.

¹ In der Studie wurden Daten für mehrere Fragestellungen erhoben. Wir beschränken uns auf die Darstellung der für die Erhebung der Usability unmittelbar relevanten Elemente. Weitere Details können bei den Autoren erfragt werden.

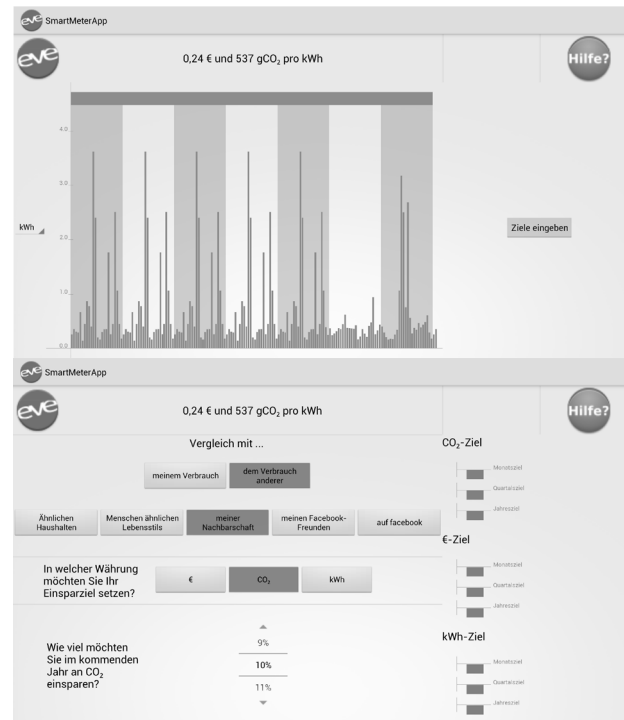


Abb.1: Screenshots der in der Studie genutzten Version des Inhome Displays

Um ein umfassendes Bild von den Probanden zu erhalten und um weitere Fragestellungen zu beantworten werden die Probanden gebeten, verschiedene Fragebögen auszufüllen. Diese umfassen unter anderem Fragen zu soziodemografischen Angaben, Umwelteinstellungen und -verhalten, Intelligenz, Persönlichkeit, Emotionen und Lebensstilen. Die Usability wurde mittels verschiedener, teilweise adaptierter, Fragebögen ermittelt: Ergonorm (Dzida et al., 2000), CSUQ (Lewis, 1993), Isonorm 9241/10 (Prümper & Anft, 1993), UEQ (Laugwitz, Schrepp, & Held, 2006), AttrakDiff (User Interface Design GmbH) und ergänzende, selbstformulierte Fragen. Um außerdem objektive Testdaten zu erhalten, wird das Gesicht der Probanden während der Interaktion mit dem Tablet abgefilmt. Diese Daten werden hinsichtlich emotionaler mimischer Reaktionen (mittels FaceReader, Trademark von Noldus IT) ausgewertet. Auch die Nutzereingaben auf dem Tablet werden registriert. Dafür wird das Videosignal des Tablets mittels HDMI Capture Card auf einem Computer eingelesen und dort in Bezug auf Reaktionszeiten und Dauer von Interaktionen ausgewertet.

Die Probanden erhalten die Aufgabe, sich zunächst mit dem Tablet und der App vertraut zu machen. Darauf folgen mehrere Aufgaben in denen sie unterschiedliche Einstellungen oder spezifische Aufgaben mit der App auf dem Tablet Computer vornehmen sollen.

Diskussion

Die erste Studie soll zeigen, ob die bisherigen Elemente von den Probanden akzeptiert werden und die Interaktion unkompliziert und angenehm ist. Während wir uns

in der vorliegenden Studie lediglich mit Aspekten des Designs beschäftigen, sind auch weitere Fragestellungen nicht zu vernachlässigen. Allen voran sollte zukünftig Augenmerk auf Fragen des Datenschutzes und der Datensicherheit gelegt werden. Während zum einen rein technische Maßnahmen die Sicherheit gegen Angriffe von außen gewährleisten müssen (Schutz gegen Hacker etc.), ist zum anderen zu klären, ob und wenn ja, wie, Nutzer von Smart Metern ihre Stromverbrauchsdaten veröffentlichen/"teilen" oder als Vergleichsdaten zur Verfügung stellen können. Risiken dieser Veröffentlichung sollten kritisch hinterfragt und analysiert werden. Weiterhin sollten die Nutzer auf diese Risiken in geeigneter Art und Weise hingewiesen werden. Dass aus detaillierten Lastgangdaten beispielsweise erkannt werden kann, wann sich Personen gewöhnlicher Weise zu Hause aufhalten und wann nicht, kann für potentielle Einbrecher eine wertvolle Information sein.

Abgesehen von offenen technischen und sicherheitsrelevanten Fragen bietet die Technologie des Smart Metering zahlreiche Möglichkeiten, den Kunden bei der Veränderung seines Stromverbrauchsverhaltens zu unterstützen. Zentral ist die Entwicklung einer Interaktionsstrategie, die auf gesicherten wissenschaftlichen Erkenntnissen beruht und empirisch auf ihre Effektivität untersucht wurde.

Danksagung

Wir möchten uns bei den Projektpartnern aus dem Institut für Hochspannungstechnik und elektrische Energieanlagen und dem Institut für Sozialwissenschaften der TU Braunschweig für die technische Umsetzung und die gemeinsamen Diskussionen bedanken.

Literatur

- Darby, S. (2006). The effectiveness of feedback on energy consumption. University of Oxford: Environmental Change Institute.
- Diekmann, A., & Preisendörfer, P. (1992). Persönliches Umweltverhalten. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 44, 226-251.
- Dzida, W., Hofmann, B., Freitag, R., Redtenbacher, W., Baggen, R., Geis, T., . . . H. Peters, H. (2000). Gebrauchstauglichkeit von Software. Ergonom: Ein Verfahren zur Konformitätsprüfung von Software auf der Grundlage von DIN EN ISO 9241 Teile 10 und 11. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin.
- European Commission. (2010, 22.09.2011). The EU climate and energy package. from http://ec.europa.eu/clima/policies/package/index_en.htm
- Franzen, A. (1997). *Umweltbewusstsein und Verkehrsverhalten*. Zürich: Verlag Rüegger.
- Gans, W., Alberini, A., & Longo, A. (2013). Smart meter devices and the effect of feedback on residential electricity consumption: Evidence from a natural

- experiment in Northern Ireland. *Energy Economics*, 36(0), 729-743. doi: <http://dx.doi.org/10.1016/j.eneco.2012.11.022>
- Hargreaves, T., Nye, M., & Burgess, J. (2010). Making energy visible: A qualitative field study of how householders interact with feedback from smart energy monitors. *Energy Policy*, 38(10), 6111-6119. doi: <http://dx.doi.org/10.1016/j.enpol.2010.05.068>
- Horn, M., Haiduk, M., & Schwarze, A. (2013). *Darstellungsformen für CO2*. Unveröffentlichtes Arbeitspapier. TU Braunschweig.
- Huber, J. (2001). *Allgemeine Umweltsoziologie*. Wiesbaden: Westdeutscher Verlag.
- Ivanov, C., Getachew, L., Fenrick, S. A., & Vittetoe, B. (in press). Enabling technologies and energy savings: The case of EnergyWise Smart Meter Pilot of Connexus Energy. *Utilities Policy*(0). doi: <http://dx.doi.org/10.1016/j.jup.2012.10.001>
- Laugwitz, B., Schrepp, M., & Held, T. (2006). UEQ - User Experience quantitativ messen. from <http://www.ueq-online.org/>
- Laurens, J. M. (2012). Changing Behavior and Environment in a Community-based Program of the Riverside Community. *Procedia - Social and Behavioral Sciences*, 36(0), 372-382. doi: <http://dx.doi.org/10.1016/j.sbspro.2012.03.041>
- Lewis, J. R. (1993). IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use (Technical Report 54.786). Boca Raton: IBM Corporation.
- Mack, B. (2007). *Energiesparen fördern durch psychologische Interventionen. Entwicklung und Evaluation einer Stromsparkampagne in einer Energiesparhaussiedlung*. Münster: Waxmann.
- Prümper, J., & Anft, M. (1993). Fragebogen ISONORM 9241/10. from http://www.ergo-online.de/site.aspx?url=html/software/verfahren_zur_beurteilung_der/fragebogen_isonorm_online.htm
- Staats, H. J., Harland, P., & Wilke, H. A. M. (2004). Effecting Durable Change: A Team Approach to Improve Environmental Behavior in the Household. *Environment and Behavior*, 36(3), 341-367. doi: 10.1177/0013916503260163
- Staats, H. J., Wit, A. P., & Midden, C. Y. H. (1996). Communicating the Greenhouse Effect to the Public: Evaluation of a Mass Media Campaign from a Social Dilemma Perspective. *Journal of Environmental Management*, 46(2), 189-203. doi: <http://dx.doi.org/10.1006/jema.1996.0015>
- User Interface Design GmbH. AttrakDiff. from <http://www.attrakdiff.de/>
- Werner, C. M., Sansone, C., & Brown, B. B. (2008). Guided group discussion and attitude change: The roles of normative and informational influence. *Journal*

of Environmental Psychology, 28(1), 27-41. doi:
<http://dx.doi.org/10.1016/j.jenvp.2007.10.002>

MErLiN – A novel multi-touch based assistance system for bio-technological laboratories

Franziska Maugg¹ and Sebastian Schöning^{1,2}

Keywords: electronic laboratory notebook, bio-technological laboratory, assistance system, multi-touch, hybrid work methodology

Abstract

We introduce MErLiN, a novel approach to on-site digital laboratory journals based on modern multi-touch technology. The concept of MErLiN uses abstracted experiment models that are specified with parameters, devices and materials when the experiment is conducted. The result is a seamless and consistent documentation. This article presents the first demonstrator of MErLiN and proposes next steps.

Introduction

The daily tasks of bio-technological laboratory technicians are not restricted to single workplaces. Lab technicians switch between their desktop, where most of the preparation and evaluation of experiments takes place, and a sterile workbench to conduct and document experiments. Since labs are mostly shared by several technicians, a strict and tidy organization of materials and labware is required. Hygiene restrictions have to be followed, especially if sterile and non-sterile routines take turns.

Experiments generally comprise a hypothesis, an experimental setup executed as a range of trials and a subsequent evaluation of results. Every phase of the experiment is documented in a lab journal that may be used for further scientific work or exploitation. This documentation has to be understandable and complete. In the simplest example, technicians document experiments in a paperback journals annotated by printed tables, diagrams, and photos. These journals bear the disadvantage that they cannot be used in sterile environments. Furthermore, the content is not digital and, therefore, not easy to share and re-use. Electronic lab notebooks (ELNs) therefore address the challenge of paper-free laboratories.

A problem of consistence emerges from semi-automated workflows. Lab devices of different complexity are commonly used to automate tasks such as pipetting a large number of samples. A parallel logging of activities is provided by most devices and can be used as seamless documentation. In reality, manual tasks are still necessary before, in between, or after automated procedures, however they aren't considered in automatically generated documentations.

¹ Fraunhofer IPA, Project Group Automation in Medicine and Bio-Technology, Mannheim

Although parts of process data exist in digital form already, aggregation and merging with documentations of manual tasks is necessary. The obvious way is to either transcribe device data manually or to import them into the lab journal.

In comparison to manual documentation whose quality and completeness rely on individual skills and habits, automatically generated documentation may show a different structure and a higher level of detail.

Related work

Science projects and commercial products propose to take the ELN functionality inside the lab to approach the problem of frequent switching between desktop computers and lab workplaces. Projects like eLabBench or BioTisch use a horizontal multi-touch capable glass-covered screen to create augmented workbenches (Tabard et al. 2011, 2012, Echtler, Häussler et al 2010). The aim is to make information available in the lab, e.g. the experiment setup description or scientific resources, and to replace physical objects like calculators and notepads with digital equivalents (Echtler et al 2010). In contrast to expensive multi-touch tables, most commercial ELN solutions offer a version for mobile devices to bring their features into the lab (e.g. LabFolder, LabArchives, and eCAT).

BioTisch guides users through processes with step-by-step instructions, whereas eLabBench offers a recipe-alike listing of tasks to be performed. In both cases, documentation is not the focus of the applications. Desktop ELNs and their mobile versions, in comparison, concentrate on the file management and text-based documentation by offering text editors and templates (Bungers 2013).

Equivalent to paperback lab journals, free text input in ELNs may result in a documentation that differs in detail and consistency. Findings in the field of text input on flat glass suggest that typing on commercial multi-touch tables is not yet sophisticated (Findlater and Wobbrock, 2012) and may even affect the way users formulate text (2011). Hence, it is reasonable to claim that sole text-editor based solutions are not the most efficient implementation in these settings.

When users assign single tasks to devices, the device output as well as the details of the automation procedure has to be included into the documentation. This may require additional modifications of the textual basis or the used template.

The required processing of automatically generated information and manual documentation shows the need of an on-site assistant system that supports a hybrid work methodology. This system should not only guide users through experiments, but also to gather information in a structured way without differing between automated and manual tasks. The aim is to minimize interactions between users and the ELN and various devices, so that the whole experiment process is more streamline.

Modeling hybrid workflows as a base for consistent experiment documentation

As a bio-medical research facility with an interdisciplinary team, the Fraunhofer PAMB accommodates a laboratory which allows to observe the typical workflows and to investigate on user requirements. Our motivation is to develop a multi-touch-based electronic lab notebook for seamless and consistent documentation of experiments. The requirements are a) on-site installation in labs, where the software interface is equivalent to the physical workbench, b) a functionality that generates structured documentations not only from fully-automated processes but also from semi-automated and manual processes and c) to integrate lab devices into the system to setup a hybrid workflow.

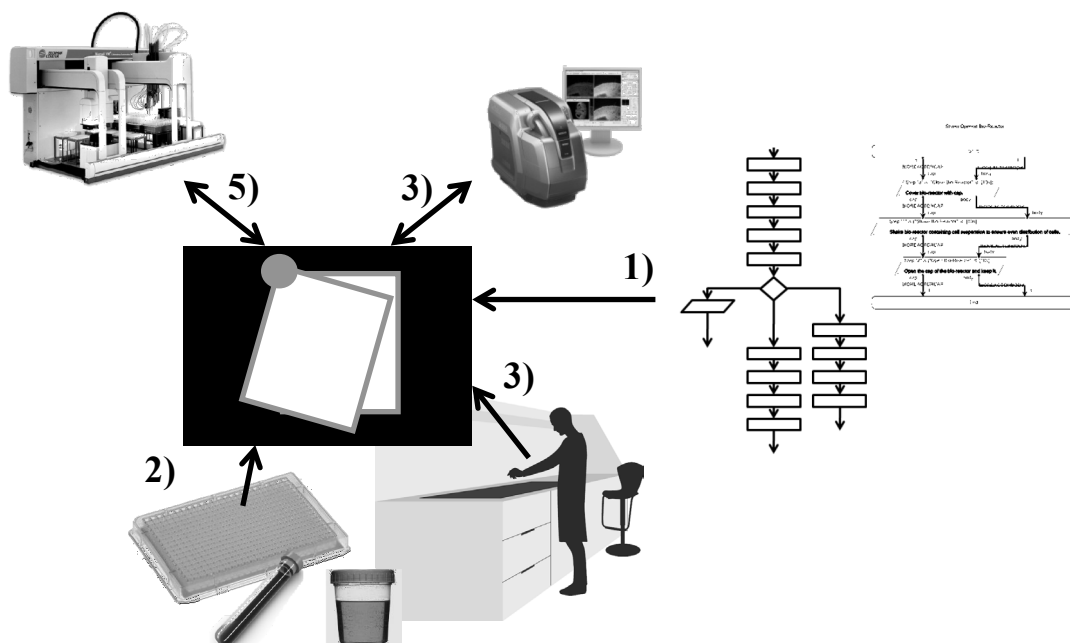


Abb. 4: Modeled experiments (1) are related to available labware and samples (2). Manual tasks are executed by users (3). Attached devices are controlled by the system whenever tasks are automated (4). Process data and measured data are retrieved automatically (5).

To achieve consistent, well-structured and re-usable results, experiment descriptions are abstracted as potentially nested material flows. Each individual step is either specified by required devices or by manual execution. Users define the equivalence between models, which enables to substitute manual steps by automated steps and vice versa.

When users conduct experiment trials, the modeled setup is related to samples and devices at hand, and concrete values of timing constraints. User document manual tasks by going through one task after another and entering measured values if required. The system proposes alternative ways to execute a step, depending on device availability. When users assign tasks to attached devices, the output data is

directly retrieved from devices by the system. Details of the device configuration are documented when the workflow is initialized.

Our modeling approach enables a seamless documentation during experiment conduct, considering not only fully-automated or exclusively-manual trials but also human-machine hybrid work flows. Abb. 4 visualizes the overall structure of this approach.

The implementation of MErLiN

The software MErLiN implements the concept presented in the previous section. A dedicated user interface is realized based on elaborate analysis of the user context and requirements (Fischbein and Röhm 2011, Pfalzgraf 2011). MErLiN provides two work modes: Firstly, a work mode that offers an experiment browser including result presentation. Secondly, a mode for experiment conducts. Experiments are displayed on the overview screen with a scalable depth of information, represented as nodes in a tangible graph with an organic look and feel. Semantic relations between the experiments are represented by graph edges. The metaphor of a paper-back lab journal is used to visualize the content. Floating screen elements can be arranged freely, such as the sidebar and little helpers like lab calculators or timers.

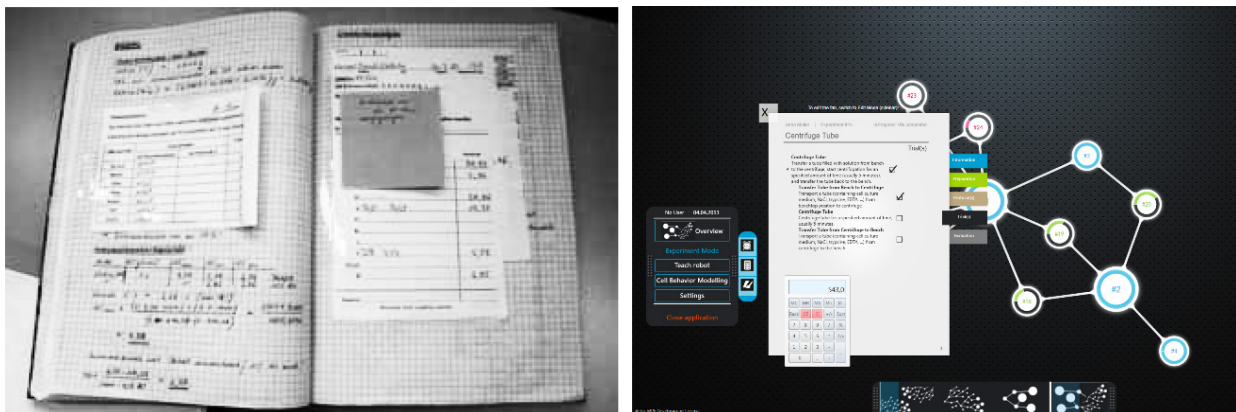


Abb. 5: A traditional laboratory journal (left) and the digital equivalent on the MErLiN user interface (right).

The journal (see Abb. 5) provides general information and notes about particular experiments and offers users to specify the experiment setup parameters. The experiment work flow is currently visualized as a check list of nested tasks. If matching devices are available, single tasks can be automated by clicking on a symbol instead of ticking off a list entry. The conduct of experimental steps is captured in the same way, either by the user or by the device.

In the context of the EU research project ManuCyte, MErLiN was installed on a multi-touch system mounted in the housing of a sterile workbench with a protective glass and laminar air flow (ManuCyte, 2013). The embedded screen is used as both software interface and working surface appropriate to safety and hygiene restrictions (e.g. operability with gloves and resistance to sterilizer). In combination

with an assisting robot that automates single tasks (RRRU, 2013), MErLiN demonstrates the hybrid working methodology.

Discussion

Compared to similar projects, MErLiN focuses on how processes of conducting and documenting experiments can be supported and captured. The current implementation allows us to evaluate the concept with respect to the specific hardware setup and use case. However, the project outcome is of general importance for further development of the software concept and user interface.

We assume that free-text editors cannot assure a documentation of homogeneous consistency and quality. The feedback of lab technicians using our prototype confirmed the usability issues of text input on flat multi-touch screens. Therefore, we implemented a concept that uses abstracted experiments to generate documentations. By this approach, input interactions can be reduced and hybrid workflows are supported. Text input, however, cannot be avoided completely and we believe that adaptive screen keyboards are a promising approach (Findlater and Wobbrock, 2012).

To get a consistent and complete documentation, the prose instructions converted from experiment models have to be followed strictly. At first glance, this seems like a lack of flexibility. Tabard et al (2012) look at molecular and synthetic biology with an exploratory and iterative work flow. They conclude that fixed instructions may prevent a creative process and therefore innovation. The MErLiN project however bases on the analysis of bio-technological laboratory procedures, where the experiment setup itself mostly remains while experiment parameters change only. We argue that creativity is not eliminated by using experiment models to support users throughout the experiment conduct. Instead, we emphasize that scientific outcome and innovation emerges from creative experiment setups enabled by flexible user interfaces. Subsequent proper conduct and well-structured documentation reflect sound scientific procedures.

Conclusion and future work

In this paper, we presented MErLiN, a novel approach to electronic lab notebooks. MErLiN is integrated inside a laboratory workbench where it serves both as physical working surface and user interface to present the information on-site where it is needed. To ensure a consistent and seamless documentation of both manually conducted, automated or human-machine 'hybrid' processes, experiments are abstracted as material flows. To execute an experiment trial, the model is projected onto available labware and presented to the user as an interactive prose instruction. Tasks intended for automation may be assigned to devices and can be executed directly.

As a next step, MErLiN is implemented in our own laboratories to support lab technicians throughout their daily tasks and to further evaluate and revise the user interface and interactions. Findings on multi-touch input on plate surfaces and the use of tabletop screens and tablets in laboratories will be considered.

The current state of MErLiN suggests investigations on process optimization in biological laboratories. Other directions of research such as novel data mining approaches are inspiring and promising in this context.

Literature

- Bungers, S. (2013). LabFolder. Retrieved July 09, 2013, from www.labfolder.com
- Echtler, F., Häussler, M., & Klinker, G. (2010). BioTISCH: The interactive molecular biology lab bench. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems - CHI EA '10* (p. 3439). New York, USA: ACM Press. doi:10.1145/1753846.1753998
- Findlater, L., & Wobbrock, J. (2012). Personalized input. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI'12* (p. 815). New York, USA: ACM Press. doi:10.1145/2207676.2208520
- Findlater, L., Wobbrock, J. O., & Wigdor, D. (2011). Typing on flat glass. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI'11* (p. 2453). New York, USA: ACM Press. doi:10.1145/1978942.1979301
- Fischbein, V., & Röhm, J. (2011). Multitouch Lab Journal. 2011. Hochschule der Medien, Stuttgart. Retrieved from <http://www.info-design.net/laborbuch/>
- ManuCyte Project. (n.d.). Retrieved July 09, 2013, from <http://www.manucyte-project.eu/>
- Pfalzgraf, Y. (2011). User Interface Prototyping und Evaluation eines elektronischen Laborbuches auf der Basis eines Multitouch Systems für die biochemischen Forschung. Hochschule der Medien Stuttgart.
- RRRU. Hochschule für angewandte Wissenschaften Regensburg (n.d.). Regensburg Robotics Research Unit (RRRU). Retrieved July 01, 2013, from <https://www.hs-regensburg.de/fakultaeten/maschinenbau/labore/robotik.html>
- Tabard, A., Hincapié Ramos, J. D., & Bardram, J. (2012). The eLabBench in the wild. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12* (pp. 3051–3060). New York, USA: ACM Press. doi:10.1145/2207676.2208718
- Tabard, A., Hincapié-Ramos, J.-D., Esbensen, M., & Bardram, J. E. (2011). The eLabBench. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces - ITS '11* (p. 202). New York, USA: ACM Press. doi:10.1145/2076354.2076391

Advanced Tactile Feedback in the Context of Sensorimotor Assistance Systems: A Tapping Task Study

Andreas Bremer¹ und Gerhard Rinkenauer¹

Keywords: tapping, surrogate feedback, tactile feedback, assistance

Abstract

Surrogate tactile feedback in the context of sensorimotor assistance systems seems to be a promising way to provide effective and meaningful sensory reafferences. An experiment was conducted to investigate the influence of additional tactile feedback on the performance of repetitive tapping tasks. Measures of variable and constant timing errors, tapping force and subjective rating of mental effort were used to assess movement performance. The results suggest that surrogate tactile feedback in healthy participants supports force control but seems to worsen, at least partially, timing control of repetitive movements. A possible reason for the detrimental effects on timing control may be due to the signal properties of the surrogate feedback or the lack of an adequate learning period. These issues are within the focus of our current research.

[Künstliches taktiles Feedback als Ersatz für natürliches Feedback im Kontext von sensorimotorischen Assistenzsystemen scheint ein vielversprechender Weg zu sein, effektiv und bedeutsam sensorische Reafferenzen anzubieten. Ein Experiment wurde durchgeführt, um den Einfluss von zusätzlichem künstlichem taktilen Feedback auf die Präzision bei einer klassischen Tapping-Aufgabe zu untersuchen. Gemessen wurden variabler und konstanter Fehler, sowie die Kraft bei den Tap-Bewegungen und die subjektive mentale Anstrengung. Die Resultate zeigen, dass zusätzliches taktiles Feedback bei gesunden Probanden die zeitliche Präzision zumindest teilweise verschlechtert, aber auch, dass zusätzliches Feedback eine ökonomischere Dosierung der Kraft beim Tappen ermöglicht. Eine mögliche Ursache für die nachteiligen Effekte auf die zeitliche Genauigkeit liegt in den Signaleigenschaften des zusätzlichen Feedbacks (verzögertes Ansprechen des magnetischen Kolbens).]

Introduction

The accurate control of movement sequencing and timing is important for many types of motor behavior (e.g. speaking, typing, walking, etc.). It is also important for a successful interaction with sensorimotor assistance systems like prostheses, orthoses, telerobot-systems. However, typical use case scenarios for such systems often comprise lack of sensory feedback. For instance, stroke or spinal cord injuries often result not only in motor, but also in sensory deficits. Sensory infor-

¹ Leibniz-Institut für Arbeitsforschung an der TU Dortmund (IfADo)

mation, nevertheless, is thought to play a crucial role for motor control since a long time (Greenwald, 1970). This underlines the notion of including surrogate feedback within such systems. While there is a lot of research activity in this area (cf. Miller & Weber, 2011), most of these approaches are somewhat impractical and/or do not convey temporal feedback aspects. Hence, a more fundamental approach might help to shed light on the mechanisms of temporal surrogate feedback for the future development of user-centered and effective systems.

Common measures of temporal precision of repetitive planned movements include constant and variable error. The variable error can further be used to localize the source of the variance according to the Wing-Kristofferson two level timing model (WKM). The WKM uses autocovariance data to divide the raw variable error into two distinct sources of variability, the central and the peripheral variance, with the first modeling the properties of a central clock like structure, while the latter models the motor implementation of the movement (Vorberg & Wing, 1996).

Generally, there is evidence that sensory reafferences contribute to the timing precision of movements. For instance, in a study by Drewing, Henning & Aschersleben (2002), the participants had to execute repetitive finger tapping movements under various conditions. Earlier studies (Helmuth & Ivry, 1996) found a bimanual tapping advantage regarding within-hand variability, which could be attributed to a lowered central variance under the assumptions of the WKM. Drewing et al. argued that the decrease in variability was due to the additional sensory reafferences present when tapping with both left and right index finger.

In another study by Stenneken, Prinz, Cole, Paillard & Aschersleben (2006), deafferented patients were compared to students and an age-matched control group at tapping tasks. The participants either tapped with additional extrinsic feedback (auditory and/or visual) or solely without any additional feedback. With full extrinsic feedback, patients were able to enhance their timing accuracy. The control group (non-deafferented), however, could not fully reduce their synchronization error. Regarding the modalities of the extrinsic feedback, it would be interesting to see whether an additional artificial feedback within the same modality (tactile) can enhance timing precision.

Therefore, to assess the potential of surrogate tactile feedback we presented additional tactile feedback to a healthy sample of participants. According to the studies mentioned above it was expected that the surrogate feedback will reduce both the central variance and constant error.

Method

Twelve right-handed participants took part in the experiment. They were drawn from a student population and had a mean age of 25.6 (SD = 4.1, range 18 - 33). Compensation for participation was 10€. All participants gave informed consent.

Previous to the actual experiment two pretests were conducted to estimate participants' general hand motor ability and sensation. The first one was a finger to thumb opposition task (adopted from Karni, Meyer, Jezzard, Adams, Turner, & Ungerleider, 1995) to ensure hand mobility. The second test was the Semmes-Weinstein-Monofilament test (SWMF), an objective sensitivity testing of the skin. The SWMF was conducted at the different sites: left and right index finger, as well as the right big toe. Both tests confirmed that subjects dispose of normal motor and tactile functioning.

The experiment consisted of a set of classical tapping tasks, which comprised a synchronization phase of 20 tones and a continuation phase of 30 tones. During the synchronization phase participants were required to synchronize their tapping movements to pacing stimuli which were delivered via headphones. The Target Inter-Tap-Intervals (Target ITIs) were either 250 or 500 ms. The pacing stimuli consisted of 1 kHz beeps with a duration of 50 ms. After the synchronization phase the participants had to keep the rhythm during the continuation phase without pacing stimuli. Each produced tapping sequence was immediately checked for large deviations (defined as an interval longer than $1.5 \times \text{Target ITI}$ or shorter than $0.5 \times \text{Target ITI}$) and repeated in case of one or more large deviations.

Tapping movement characteristics were measured by a custom designed device. The timing of the movements was acquired by light barriers (mounted about 3 millimeters above the surface), whereas a force key was used to measure the exact force applied during the tapping movements. Data from the light barrier and the force key were recorded by an A/D-Converter (Labjack U12). The experiment was controlled by a Python (version 2.7.3) program.

During the tapping task, participants had to place their right foot on a box (about 14 cm high). The box was equipped with a magnetic plunger, located beneath the right big toe. The top of the magnetic plunger was modified with a rounded plastic knob. The magnet was active in half of the trials of the experiment (Feedback conditions). Participants received no advanced information about when the magnet was active. The magnetic plunger moved upward until it knocked the toe's skin and was triggered by the onset of the force key signal. The delay between force signal onset and skin contact was about 12 ms. Order of Feedback condition and Target ITI was based on a Latin square pseudo-randomization. The complete design was repeated in four blocks (factor Block).

Statistical analyses were performed using the statistical computing environment R (R Development Core Team, 2013). Primary dependent variable of interest was the variability of the produced ITIs, which was split up according to Wing and Kristofferson (1973a, 1973b), to compute the estimates for the central and peripheral variance component, which were in turn assessed by a $\text{Feedback}(2) \times \text{Target ITI}(2) \times \text{Block}(4)$ ANOVA. Furthermore, the constant error, defined as the deviation of the mean produced ITIs from the Target ITIs, was determined. Additionally, the subjective mental workload was assessed by administering the Rating

Scale of Mental Effort (RSME; Zijlstra, 1993) for each condition. In case of sphericity violations, Greenhouse-Geisser correction was applied. Generalized eta squared (η^2_G) was computed as suggested by Olejnik and Algina (2003).

Results

The central variance component showed a significant effect of Target ITI ($F(1,11) = 77.71$, $p < .001$, $\eta^2_G = .403$). The longer intervals (500 ms) elicited a higher mean central variance ($M = 764.2 \text{ ms}^2$, $SD = 1269.2 \text{ ms}^2$) than the faster (250 ms) intervals ($M = 254.9 \text{ ms}^2$, $SD = 433.9 \text{ ms}^2$; see Fig. 1a). There were no further differences found within the central variance component.

The peripheral variance in turn was influenced by the additional feedback ($F(1,11) = 7.8$, $p = .017$, $\eta^2_G = .036$). When the magnet was active, the peripheral variance ($M = 76.5 \text{ ms}^2$, $SD = 385.9 \text{ ms}^2$) was significantly higher than during tapping without additional feedback ($M = 41.8 \text{ ms}^2$, $SD = 325.4 \text{ ms}^2$; see Fig. 1b).

Furthermore, an interaction effect of Feedback and Target ITI on the constant error was found ($F(1,11) = 7.96$, $p = .017$, $\eta^2_G = .011$). Without feedback, participants generally produced a higher constant error during the longer Target ITIs ($M = -13.28 \text{ ms}$, $SD = 25.09 \text{ ms}$) than during the shorter target ITIs ($M = -5.34 \text{ ms}$, $SD = 22.9 \text{ ms}$; $p < .001$). Additionally, with feedback, the constant error was found to be larger during the faster (250 ms) Target ITI ($M = -11.03 \text{ ms}$, $SD = 24.13 \text{ ms}$) than without feedback ($M = -5.34 \text{ ms}$, $SD = 22.9 \text{ ms}$, $p = .008$). Instead, when tapping to longer (500 ms) Target ITIs, the constant error was larger without feedback ($M = -13.28 \text{ ms}$, $SD = 24.09 \text{ ms}$) than with feedback ($M = -8.6 \text{ ms}$, $SD = 28.93$; $p = .049$; see Fig. 1c).

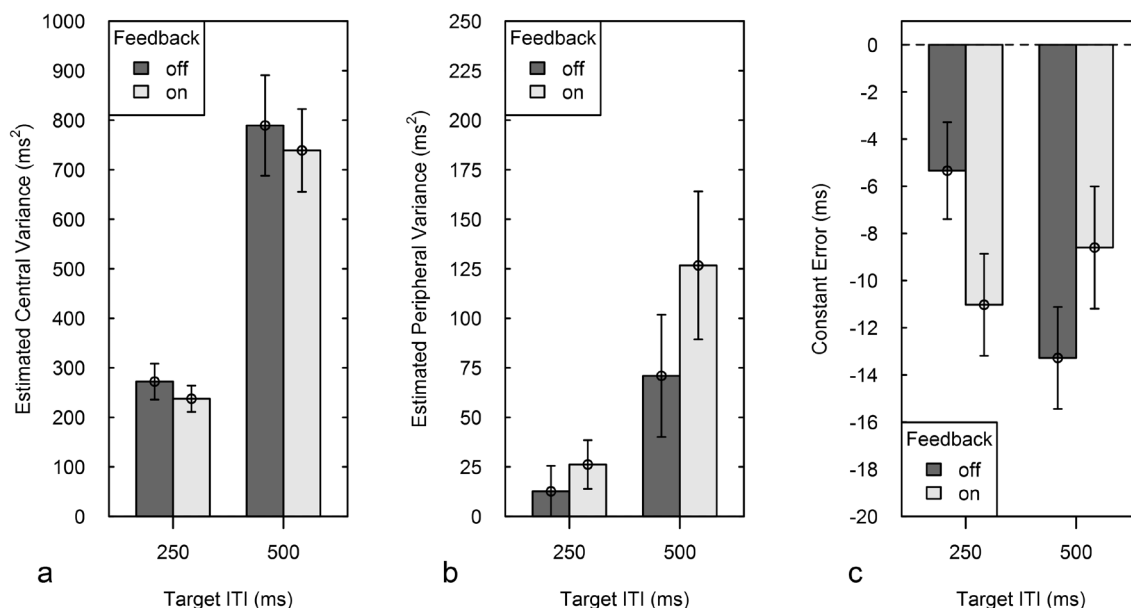


Figure 1: a) and b) central variance resp. peripheral variance estimated from the tapping experiments according to the Wing-Kristofferson model. c) constant error (deviation of the

produced interval duration from the Target ITI). Error bars indicate 95-percent confidence intervals.

The subjective ratings of mental effort as measured by the RSME also revealed a main effect of Feedback ($F(1,23) = 4.56$, $p = .044$, $\eta^2_G = .005$): participants rated tapping with the magnet on ($M = 23.26$, $SD = 14.06$) higher than when there was no additional feedback ($M = 21.19$, $SD = 11.92$; Abb. 2a). There was also a marginal interaction effect of Block and Target ITI ($F(1.6,37.5) = 3.28$, $p = .053$, $\eta^2_G = .007$; Abb. 2b) which just failed to be significant. Tukey post-hoc tests showed that tapping during longer Target ITIs was rated higher at the last two blocks (Block 3: $M = 25.57$, $SD = 10.62$; Block 4: $M = 25.09$, $SD = 9.12$) than at the first block ($M = 17.86$, $SD = 9.74$; $p = .001$ resp. $p = .004$).

Tapping force analysis also yielded a significant difference as a function of Feedback ($F(1,11) = 5.94$, $p = .033$, $\eta^2_G = .01$). While the magnet was active, participants applied less force during their tapping movements ($M = 56.55$ cN, $SD = 77.12$) compared to when feedback was off ($M = 75.45$ cN, $SD = 110.19$; Fig. 2c).

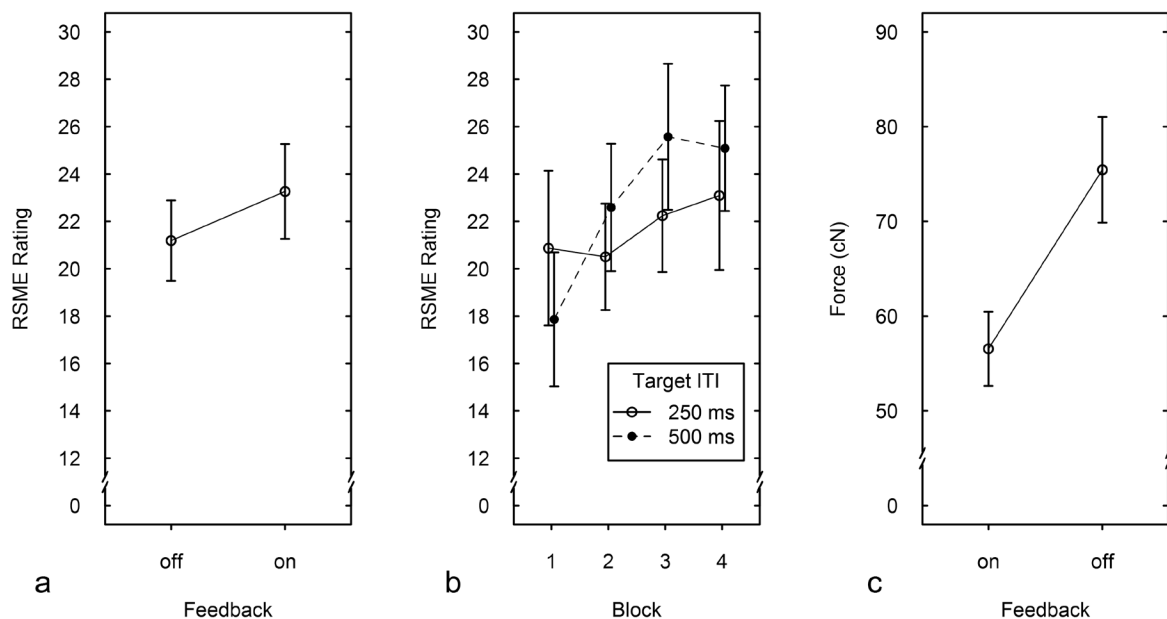


Figure 2: a) RSME rating across Feedback. b) RSME rating over all blocks, split by Target ITI. c) Mean force exerted during tapping movements. Error bars indicate 95-percent confidence intervals.

Discussion

The present study's aim was to assess whether an additional, surrogate tactile feedback apart from the acting effector can enhance timing precision of repetitive movements. The results, however, were mixed. First of all, the multi-effector advantage found by Drewing et al. (2002), i.e. lowering the central variance, could not be replicated. Instead, contrary to our expectations, peripheral variance was

increased under additional feedback conditions. It may be the case that the systematic activation delay of the magnet was responsible for these patterns. A new feedback device, with a negligible delay, has been developed and is currently being deployed to test this hypothesis.

Considering the constant error, Feedback worsened timing precision during the faster target ITI, but increased timing precision during the slower Target ITI. Presumably, there was an interference with the afore-mentioned 12 ms onset delay and the fast Target ITI, which participants could not handle, while during the slower Target ITI, participants could fully take advantage of the additional sensory reafferences. Testing this assumption will be subject of a follow-up study.

Interesting is also the finding that the additional feedback reduces the applied force during tapping movements. Generally, force regulation during movements is aimed at maximizing the economy of the movement, hence, using as little power as necessary. For instance, when small objects are lifted, grip force is regulated in a way that the object is close to slipping through the fingers to prevent injuries, damage or muscle fatigue (Johansson & Westling, 1984). In this context, the feedback can be seen as improving the economy of tapping movements. Further research is required to see whether this effect persists and where it originates.

Regarding the field of sensorimotor assistance systems, the findings of the present study offer interesting insights. First, additional feedback seems to support the economic regulation of exerted force. It seems to be a plausible assumption that such benefits are transferable to surrogate feedback concepts for future assistance systems. Second, there were positive effects of feedback on the constant error under certain conditions. However, there were also detrimental effects of Feedback on timing precision. Thus, uncovering the mechanics beyond the sensorimotor integration of surrogate feedback remains a challenge, but, given the possible benefit, is definitely worth the effort.

Literature

- Drewing, K., Hennings, M., & Aschersleben, G. (2002). The contribution of tactile reafference to temporal regularity during bimanual fingertapping. *Psychological Research*, 66, 60-70.
- Greenwald, A. G. (1970). Sensory feedback mechanism in performance control: With special reference to the ideo-motor mechanism. *Psychological Review*, 77(2), 73-99.
- Helmuth, L. L., & Ivry, R. B. (1996). When two hands are better than one: reduced timing variability during bimanual movements. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 278-293.
- Johansson, R. S., & Westling, G. (1984). Roles of glabrous skin receptors and sensorimotor memory in automatic control of precision grip when lifting rougher or more slippery objects. *Experimental Brain Research*, 56, 550-564.

- Karni, A., Meyer, G., Jezard, P., Adams, M. M., Turner, R., Ungerleider, L. G. (1995). Functional MRI evidence for adult motor cortex plasticity during motor skill learning. *Nature* 377, 155-158.
- Miller, L. E., & Weber, D. J. (2011). Brain training: cortical plasticity and afferent feedback in brain-machine interface systems. *IEEE Transactions on neural systems and rehabilitation engineering*, 19(5), 465-467.
- Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, 8(4), 434-447.
- R Development Core Team. (2013). *R: A language for data analysis and graphics*. Vienna, Austria.
- Stenneken, P., Prinz, W., Cole, J., Paillard, J., & Aschersleben, G. (2006). *Brain research*, 1084, 123-131.
- Vorberg, D., & Wing, A. M. (1996). Modeling variability and dependence in timing. *Handbook of perception and action*, 2, 181-262. Amsterdam, NL: Elsevier.
- Wing, A. M., & Kristofferson, A. B. (1973a). The timing of interresponse intervals. *Attention, Perception, & Psychophysics*. 13(3), 455-460.
- Wing, A. M., & Kristofferson, A. B. (1973b). Response delays and the timing of discrete motor responses. *Attention, Perception, & Psychophysics*. 14(1), 5-12.
- Zijlstra, F. (1993). *Efficiency in work behaviour: A design approach for modern tools*. Unpublished doctoral dissertation, Technical University, Delft, The Netherlands.

Role-Based-Client Workspace – Entwicklung von Dashboard-Interfaces im Product Lifecycle Management (PLM)

Ralph Tille¹, Michael Burmester¹ und Katharina Schippert¹

Schlüsselwörter: Dashboard, Role-Based Interface, Product Lifecycle Management, PLM

Zusammenfassung

Moderne Produktentstehungsprozesse erfordern auch auf Softwareseite einen effektiven und effizienten Umgang mit den Nutzeranforderungen. Für wiederkehrende, komplexe und hoch dynamische Aufgaben empfiehlt sich die Entwicklung eines rollenspezifischen Interfaces als Information-Dashboard (Role-Based Client). Im Rahmen der vorliegenden Studie wurde das Interface-Konzept mit 6 Teamleitern eines Automobilzulieferers im Rahmen einer qualitativen Nutzerstudie untersucht. Gerade der hohe Grad Individualisierbarkeit des Interface für unterschiedliche Arbeitsstile und das flache Navigationskonzept für komplexe Aufgabenstrukturen werden durchwegs sehr positiv bewertet.

Einleitung

Effiziente Produktentstehungsprozesse sind im Automobilzulieferbereich wesentliche Faktoren für eine schnelle, kostengünstige und qualitativ hochwertige Produktentwicklung. Um die vielfältigen und komplexen Prozesse überschauen und kontrollieren zu können, ist ein maßgeschneidertes PLM-Konzept (Product Lifecycle Management) in heutigen Unternehmen mit vielen Mitarbeitern und deren unterschiedlichen Rollen und Aufgaben verpflichtend (Stark, 2005; Sendler, 2009).

Die Komplexität und Diversität der Aufgaben bei hohem Zeitdruck kann für die Mitarbeiter zu Problemen, wie Verzögerungen und Fehler beim schnellen Finden, Auswählen, Navigieren oder in der Darstellung der relevanten Informationen führen (Gundelsweiler & Reiterer, 2008). Ristic, Rakic-Skokovic und Ali (2011) zeigen mit Philpotts (1996), dass es mehrere Möglichkeiten gibt, PLM-Systeme an die Nutzerbedürfnisse anzupassen. Die Rolle eines Nutzers im Prozess definiert seine Aufgaben, die in den Software-Features abgebildet sein müssen (Findlater, McGrenere & Modjeska, 2008). Die hier dargestellte Studie widmet sich der empirischen Validierung einer rollenbasierten Dashboard-Benutzungsoberfläche, die in einem Vorprojekt durch ein partizipatives Verfahren entwickelt wurde (Tille & Burmester, 2013).

Referenzen und Vorüberlegungen zu PLM und Information-Dashboards

Nicht nur in Großunternehmen spricht man heute vom Produkt-Lebenszyklus-Management, kurz PLM. Damit meint man zunächst ein strategisches und organisatorisches Konzept, um alle Erfordernisse eines Produktes von der Idee über die Entwicklung und Entstehung bis hin zu Service und Recycling abzudecken und zu kontrollieren (vgl. Sandler, 2009). Sämtliche relevanten Informationsbestände für die Produktlebenszeit gilt es, optimal zu integrieren und verfügbar zu machen. Ein wesentlicher Bestandteil ist das Produkt-Daten-Management (PDM), also die umfassende und nahtlose Verwaltung sämtlicher produktrelevanter Daten. In der vorliegenden Studie hat dies hohe Relevanz, da mehrere Nutzergruppen über einen Role-Based Client auf sehr viele Unternehmensdaten zugreifen können müssen. Die Tätigkeiten unserer Primärzielgruppe kaufmännisch-technischer Teamleiter erstrecken sich hauptsächlich über die Bereiche der Produktplanung und -definition, Entwicklung und Konstruktion sowie der Produktionsvorbereitung bis hin zum Serienanlauf (vgl. Ehrlenspiel 2007, S.158).

Bisherige Lösungsansätze in der Forschung zu PLM-Systemen beschäftigen sich intensiv mit den zur Bewältigung der anfallenden Aufgaben notwendigen technischen Voraussetzungen, Prozessen und Architekturen (z.B. Abramovici & Schulte, 2006; Merminod & Blanco, 2008). Das Role-Based Interface kann als eine Komponente für das Projekt- und Aufgabenmanagement im Produkterstellungsprozess angesehen werden. Vergleichbar eines Management-Information Systems (MIS) sollten die Informationen relevant, verfügbar und aktuell sein sowie den folgenden Qualitätsanforderungen entsprechen (Gupta 2011, S. 13):

- Eine der Situation angemessene Genauigkeit
- Eine passende, nutzerzentrierte Darstellungsform
- Hohe Verfügbarkeit, Aktualität und Relevanz

In dieser Studie soll zentral die Frage, wie eine effektive und effiziente Benutzungsschnittstelle eines Role-Based Client gestaltet sein muss, bearbeitet werden. Den Nutzern sollen aufgabenangemessen relevante Informationen im Überblick dargestellt und für Detailbearbeitungen der Zugang zu verschiedenen Programmen ermöglicht werden. Gestaltungshinweise zu Information-Dashboards finden sich bei Few (2006) oder Rasmussen, Chen und Bansal (2009). Dashboard-Interfaces setzen sich – ähnlich wie bei einem Fahrzeug-Cockpit – aus einzelnen Komponenten zusammen, und bieten dem Nutzer die wichtigsten Statusinformationen an einer zentralen Stelle (vgl. Shamsuzzoha, Ferreira, Azevedo, Faria & Helo, 2013). Velcu-Laitinen und Yigitbasioglu (2012, S. 41 ff.) sehen Dashboards auch als schnelles Überblicks- und Diagnosewerkzeug für Manager, merken aber auch an, dass noch wenig empirische Ergebnisse über die Wirksamkeit der vorgeschlagenen Gestaltungshinweise zu finden sind. Few führt weiter aus, dass ein Information-Dashboard an die speziellen Anforderungen der jeweiligen Nutzergruppe und de-

ren Aufgaben angepasst sein soll (2006, S. 36). Auch hier wird auf Individualisierbarkeit hingewiesen und liefert für uns die gedankliche Klammer zu einem angepassten, rollenbasierten Dashboard-Interface. Grundsätzliche Anforderungen für Dashboard-Interfaces sind nach Few (2006):

- leichte, einfache Überschaubarkeit,
- wenig bis keine Ablenkung,
- Wichtiges soll schnell auffindbar sein,
- Daten (Diagramme) sollen schnell auswertbar sein.

Abschließen lässt sich das mit einer Definition ebenfalls bei Few (2006, S.34):

„A dashboard is a visual display of the most important needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance.“

Ausgangssituation

Die Bedürfnisse und Abläufe von technisch-kaufmännischen Teamleitern eines großen Automobilzulieferers wurden in einer Nutzungskontextstudie anhand eines Methoden-Mix erhoben und als Komponenten in einem Benutzungs- und Rollenmodell hinterlegt (Tille & Burmester, 2013). Ein Benutzungsschnittstellenkonzept muss demnach hochgradig flexibel anpassbar sein, da bspw. der Projekttyp, die Phase, die individuelle Arbeitsweise und der Einsatz der Software-Anwendungen der technisch-kaufmännischen Teamleiter sehr heterogen sind. Gemäß dem Mantra von Shneiderman (1996) „Overview first, zoom and filter, then details-on-demand“ sollen Informationen zu mehreren Projekten, Aufgaben und Werkzeugen schnell und einfach ausgewählt, dargestellt, gefiltert und vergrößert werden, bevor man in die entsprechenden Software-Werkzeuge wie bspw. Kalkulation wechselt. Anpassbarkeit wird durch in Inhalt und Anzahl frei konfigurierbare Arbeitsbereiche erreicht. Diese werden als Workspaces bezeichnet und können individuell mit Aggregationen von Datenwerten aus bestimmten Programmen bestückt werden. In Abbildung 1 sind die wesentlichen Konzeptbestandteile dargestellt. Es können mehrere Workspaces angelegt werden, welche die Nutzer mit verschiedenen Vorschaumöglichkeiten ausstatten können. Die aktuelle Programmvorschau ist in drei Zoomstufen verfügbar. Die Arbeitsoberfläche der Workspaces lässt sich als Template anlegen und verwalten, somit kann den Nutzern eine Arbeitsvorlage zur Verfügung gestellt werden, die individuell an die zu bewältigenden Aufgaben im Arbeitsprozess angepasst werden kann.

Der Nutzer kann sich Workspaces frei anlegen und konfigurieren, um unterschiedliche Arbeitsweisen zu unterstützen und die Workspaces für verschiedene Projektphasen und Aufgaben anzulegen. Zwischen allen angelegten Workspaces kann der Nutzer frei wechseln. Über das Feld mit dem Zeichen „+“ kann ein neuer Workspace angelegt werden. Der Nutzer bekommt während der Konfiguration alle Vorschaumöglichkeiten die ihm zur Verfügung stehen angezeigt (siehe Abb. 2).

Dies bietet einen hohen Grad an Flexibilität. Je nach Nutzergruppensystematik werden entsprechende Programmangebote dargestellt. So kann der Workspace die unterschiedlichen Anforderungen erfüllen. Die Programmvorschau bildet ein Hauptmerkmal des Konzepts. Über eine einfache Drag'n'Drop-Interaktion kann der Nutzer die Vorschaukacheln auf die gewünschte Position in seinem Workspace ziehen. Die Vorschaukacheln können frei innerhalb eines Rasters positioniert werden. Dieses Raster gibt die Einschränkung vor, dass sich keine Informationselemente überlappen können, so behält der Nutzer alles im Überblick (vgl. das Motto: overview first (...) details on demand).

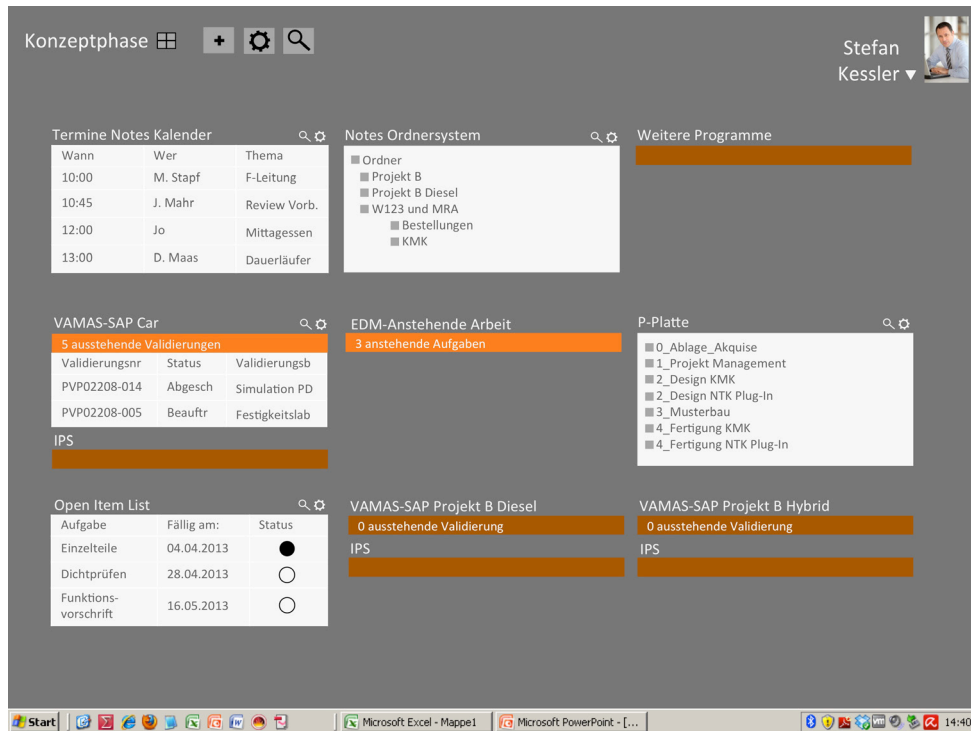


Abbildung 1: Basislayout eines Workspaces

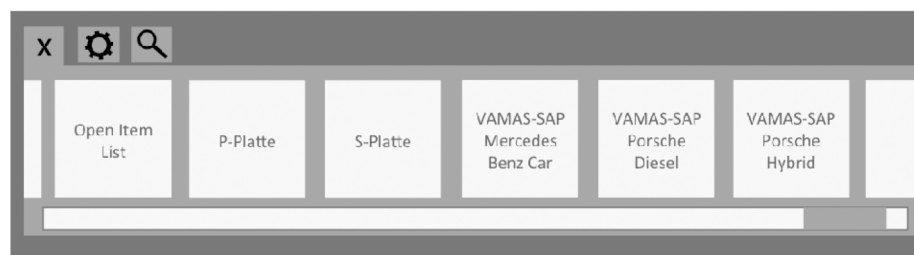


Abbildung 2: Auswahl der Programmvorschau

Die Programme, die der Nutzer auf seinem Workspace platziert, werden zu Beginn in der mittleren Vorschaugröße angezeigt. Die Vorschaugröße kann dem Informationsbedürfnis angepasst werden. Das Prinzip der 3 Vorschaugrößen (Abb.3 und 4) ermöglicht den Nutzern die Darstellung unterschiedlicher Detailgrade. Dies ermöglicht die Informationsdarstellung unterschiedlicher Komplexität vor Aufruf der Programme. Über die Pfeilspitzen kann der Nutzer zwischen den Vorschaugrößen wechseln. In der größten Detaildarstellung sind alle verfügbaren Informationen

des Programms dargestellt und scrollbar. Die anderen Programme werden dann automatisch in der kleinsten Vorschaugröße dargestellt und bleiben somit im Sichtfeld des Nutzers.

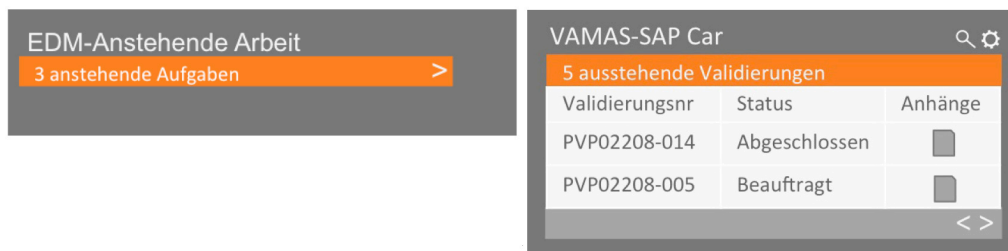


Abbildung 3: Vorschau klein und mittel

VAMAS-SAP Car								
5 ausstehende Validierungen								
Validierungs...	Status	Anhänge	No	Validierungsbereich	Validierungsart	Phase	Prüfparameter	Anzahl
PVP02208-014	Abgeschlossener			FESTIGKEITSLABOR	SENSORAPPLIK	DV	<input type="checkbox"/>	
PVP02208-002	Abgeschlossener				TEMPERATURW	DV	<input type="checkbox"/>	
PVP02208-017	Abgeschlossener				INNENDRUCKW	DV	<input type="checkbox"/>	
PVP02208-005	Beauftragt				WECHSEL TEST	DV	<input type="checkbox"/>	
PVP02208-026	Beauftragt					DV	<input type="checkbox"/>	
PVP02208-025	Dokumentation					DV	<input type="checkbox"/>	
PVP02208-031	Dokumentation					DV	<input type="checkbox"/>	

Abbildung 4: Vorschau groß

Ziele, Fragestellung und Methodik

Das Hauptziel der vorliegenden Studie ist die Validierung des zuvor in einem partizipativen Verfahren entwickelten Konzeptentwurfs (Tille & Burmester, 2013) mit denselben Teamleitern der Vorstudie. Das Interface soll zukünftig ziel-, aufgaben- und effizienzorientiert sein, d.h. es soll den Nutzern für spezifische Aufgaben die Informationen direkt, schnell und ohne Umwege bereitstellen, um eine effizientere Arbeit zu gewährleisten.

Im vorangehenden Projekt wurde eine Persona entwickelt, für die ein Aktivitäts-Szenario erstellt wurde, welches typische Abläufe im Projektalltag beinhaltet. Nach dem Vorgehen wurde für dieses Szenario ein Interface- und Interaktionskonzept entworfen, dass in Form eines Szenarioprototyps bzw. Mock-Ups umgesetzt wurde (vgl. Scenario-Based Design nach Rosson & Carroll, 2002). Das Interface wurde bewusst als hochwertige, aber nicht zu perfekt wirkende Wireframe-Darstellung umgesetzt, um die Nutzer zu Anmerkungen und Änderungswünschen zu ermutigen und gleichzeitig sicherzustellen, dass die Nutzer sich in das Anwendungsszenario hineinversetzen können. Das Szenario stellt eine bekannte Arbeitssituation dar. Es beschreibt typische Aufgaben, die ein Teamleiter mit dem Workspace erledigen kann. Die Teamleiter sollen sich mit den beschriebenen Aufgaben identifizieren, um leichter ein Urteil zu treffen, wie die verschiedenen Funk-

tionen des Workspace die Aufgabenbearbeitung in verschiedenen Arbeitssituationen unterstützt. Die Hauptfragestellungen der Studie sind:

- Können die vorgestellten Aufgaben des Szenarios mit dem Role-Based Dashboard-Interface bewältigt werden?
- Sind die Konzeptstrategien des Interfaces verständlich und nützlich?
- Entspricht das Interface den Nutzerbedürfnissen?

Das gesamte Vorgehen der Studie lässt sich wie folgt darstellen:

- Szenario- und Prototypmodellierung auf Basis der vorhergehenden Studie,
- Validierung aus Nutzersicht durch „Cognitive Szenario Walkthrough“,
- Beschreibung und Priorisierung der Anforderung auf Basis der Validierung,
- Anforderungenerstellung als Teil eines Lastenhefts für Software-Entwickler.

Die Studie wurde auf Basis der „Cognitive Walkthrough With Users“ Methode durchgeführt und anschließend qualitativ ausgewertet. Die Sitzungen wurden auf Video aufgenommen und nachträglich protokolliert. Die Protokolle bildeten die Grundlage für die qualitative Analyse. Die Analyseergebnisse wurden zu Themenfeldern aggregiert und nach ihrer Wichtigkeit priorisiert (hoch, mittel, gering). Die Wichtigkeit unterteilt sich in die Bereiche Nutzerbewertung und Umsetzungsrelevanz aus den praktischen und theoretischen Anforderungen.

Cognitive-Walkthrough-Methoden werden im HCI-Bereich sehr umfangreich, vielseitig und erkenntnisreich eingesetzt, aber auch kontrovers diskutiert. Eine Übersicht und Einschätzung sehr vieler Methoden liefern bspw. Mahatody, Sagar und Kolski (2010). In unserer Studie verwendeten wir als Basismethoden den grundsätzlichen Ablauf des „Cognitive Walkthroughs With Users“ nach Granollers & Lorés (2006) sowie die „Scenario-Based Walkthrough-Methode“ nach Segawa Sugimura, und Ishigaki (2005). Dies kann als eine Verfeinerung des „Heuristic Walkthroughs“ von Sears (1997) gesehen werden.

Der „Cognitive Walkthrough With Users“ integriert die Nutzer in den Ablauf, d.h. der Interviewer geht in unserem Fall gemeinsam mit denselben 6 Teamleitern aus der Vorstudie durch das Szenario bzw. das Interface. Nach einer kurzen Vorstellung und Erklärung der Prinzipien und Hintergründe unseres Interface-Konzeptes wird das Szenario mit den enthaltenen Aufgaben und Zielen der Persona durch die Interviewer vorgestellt und mit den Teilnehmern Schritt für Schritt durchgegangen. Währenddessen sind die Teilnehmer angehalten ihre Gedanken, Empfindungen und Meinungen laut zu äußern („Think-Aloud“-Methode). Am Ende wird mit den Teilnehmern nochmals auf Besonderheiten zu bestimmten Stellen eingegangen und eine abschließende Wertung erstellt.

Im Gegensatz zum beschriebenen Verfahren des „Scenario-Based Walkthroughs“ von Segawa et al. (2005), bei dem ohne Nutzerbeteiligung ein Szenario evaluiert wird, verwendeten wir ein Szenario als Basiselement, um die Kon-

zeptideen verständlich anhand von typischen Aufgabenabläufen darzustellen. Dieses Szenario entwickelten wir auf Basis der Ergebnisse der Vorstudie, die uns einen präzisen Einblick in den Projekt- und Tagesablauf der Zielgruppe erlaubte.

Aufgrund der Tatsache, dass das Interface in dieser frühen Phase lediglich als Mock-Up bzw. Clickdummy zur Verfügung stand und daher nicht selbsttätig exploriert werden konnte, entschieden wir uns für die Think-Aloud Methode, um die Nutzerreaktionen während bzw. nach Vorstellung einzelner Etappen innerhalb des Szenarios zu erhalten. Die bekannten Nachteile dieser Methode (bspw. Russo et al., 1989), wie z.B. zusätzliche Belastung der Nutzer oder sozial erwünschte Antworten wurden beim Studiendesign in Betracht gezogen, mit dem Ziel, solche Situationen rechtzeitig durch sensible, erfahrene Interviewer zu erkennen und entsprechend zu reagieren.

Ergebnisdarstellung

Das Konzept wurde von den 6 Teamleitern (5 männlich, 1 weiblich, alle langjährige Teamleitungserfahrung) durchgehend positiv bewertet. Die flexible Gestaltung der Workspaces unterstützt die Anforderungen nach nutzerzentrierter, aufgabenspezifischer Darstellung und Klarheit bzw. Übersichtlichkeit. Die unterschiedlichen Zoomstufen ergänzen dies durch leichte und einfache Überschaubarkeit für eine schnelle Status-Diagnose. Die qualitative Analyse der Äußerungen ergab insgesamt 14 Themenfelder. In der Beschreibung der Themenfelder wurden jeweils dargestellt die nicht funktionalen Anforderungen (Eigenschaften, die das System haben soll) und die funktionalen Anforderungen (was das System tun soll). Die Themenfelder wurden durch 3 Usability Experten diskutiert nach Nutzerbewertung und Umsetzungsrelevanz in 3 Abstufungen (hoch, mittel, gering) priorisiert.

Die Teamleiter bearbeiten meist gleiche Aufgabentypen jeweils für verschiedene Projekte oder unterschiedliche Projektphasen. Hohe Priorisierung erreichten daher die flexibel konfigurierbaren Workspaces, die von den 6 Teamleitern sehr positiv eingeschätzt wurden und eine Hauptfunktionalität des Gesamtkonzepts bilden. Die aktuelle Vorschau der verschiedenen Programme verspricht ebenso einen Effizienzvorteil und wurde mit Priorität „hoch“ eingestuft. Teilnehmer 3 sagte dazu: „Gut, weil jeder die Spaces sowohl projektbezogen als auch aufgabenbezogen anlegen kann, das halte ich für einen großen Vorteil.“ Nur auf Grundlage aktueller Informationen sind die Teamleiter in der Lage ein Urteil darüber zu fällen, ob der Absprung in das Programm erfolgen muss, oder zu diesem Zeitpunkt nicht notwendig ist. Ebenfalls „hoch“ priorisiert wurden die Möglichkeit der Größenänderung und die Vorschaukonfiguration. 4 der 6 Teilnehmer äußerten sich positiv über diese Funktion. Teilnehmer 4 meinte dazu: „Ich kann zwischen den Größen wechseln ohne groß hin- und herspringen zu müssen, das gefällt mir gut.“ Durch die Größenänderung können die Teamleiter entscheiden, welchen Detailgrad an Informationen sie sich anzeigen lassen. Diese Flexibilität gewährleistet einen Überblick der verschiedenen Programme sowie die Möglichkeit einzelne Programme im

Detail einzusehen, ohne dass der Aufruf des dahinterliegenden Programms erfolgt. Über die Konfigurierbarkeit der Vorschau können die Teamleiter auswählen, welche Informationen des Zielprogramms für sie relevant sind und eine personalisierte Vorschau erstellen.

Die weiteren Themenfelder des Workspace wurden niedriger priorisiert und sind im Folgenden aufgelistet. In Klammern ist die Anzahl der Teilnehmer gegeben, die sich positiv über die Funktion äußerten.

Priorisierung mittel:

- Single Sign on, einmaliges Einloggen für alle Programme (6 von 6)
- Starten der letzten Sitzung (6 von 6)
- Wichtige einzeilige Statusmeldungen aus den Programmen (6 von 6)
- Mouseover, Anzeige anstehender Aufgaben aus den Programmen (5 von 6)
- Sortierbarkeit der Programmvorschautabellen (6 von 6)
- Programmspezifische Suchmöglichkeit (6 von 6)
- Priorisierung gering:
- Einsehen anderer Workspaces im Falle von Projektvertretungen (6 von 6)
- Einblenden allgemeiner Systemmeldungen (6 von 6)

Die Themenfelder mit der Priorisierung mittel und gering wurden zwar fast durchwegs positiv bewertet, waren jedoch für die tägliche Aufgabenbearbeitung weniger wichtig, was dann zu einer niedrigeren Priorisierung führte. Insgesamt äußerten die Teilnehmer kaum negative Kommentare. 4 Aspekte wurden als verbesserungswürdig bewertet. 1 Teilnehmer empfand die Darstellung der Detailansicht als zu klein und würde eine Ansicht in voller Screengröße bevorzugen. Die anderen Kommentare bezogen sich auf den konkreten Inhalt der verschiedenen Programmvorschauen.

Insgesamt wurde die flache Hierarchie, welche die Kernphilosophie des Konzepts bildet, positiv hervorgehoben wie auch die einfache Interaktion und flexible Konfigurierbarkeit. Das Konzept unterstützt die Teamleiter bei ihren unterschiedlichen Tätigkeiten. Alle Teamleiter bekundeten Interesse, das Konzept in ihrem Arbeitsalltag auszuprobieren. Das Informationsbedürfnis der Teamleiter, im Zuge ihrer verschiedenen Tätigkeiten im Bereiche Produktplanung und –definition, Entwicklung und Konstruktion sowie der Produktionsvorbereitung bis hin zum Serienanlauf, wird nach Meinung der Teamleiter über einen Role-Based Client dieser Form abgedeckt.

Ausblick

Das entwickelte Konzept unterstützt die Zielgruppe in ihrer täglichen Aufgabenbearbeitung und soll als lauffähiger Prototyp für den Alltagseinsatz entwickelt werden. Die Validierung durch die Teamleiter zeigt einen hohen Grad an Akzeptanz.

Die Teamleiter äußerten den Wunsch, das System in dieser Form im Arbeitsalltag auszuprobieren. Schwierigkeit der Konzeptumsetzung sind eingeschränkte oder nicht vorhandene Systemschnittstellenzugänge sowie inhomogene Datenbasen. Sofern weitere Studien positiv ausfallen, kann das Konzept auf andere Rollen übertragen und als generelles Interaction Pattern beschrieben werden.

Literatur

- Abramovici, M. & Schulte, S. (2006). PLM – State of the Art and Trends. In: Schützer, K. (Hrsg.): *Inovacoes Tecnologicas no Desenvolvimento do Produto*, 11. Internationales Hochtechnologie Symposium „Technologische Innovationen Produktentwicklung“, 05.10.2006, Piracicaba, Brasilien.
- Ehrlenspiel, K. (2007). Integrierte Produktentwicklung. Denkabläufe, Methodeneinsatz, Zusammenarbeit. 3. akt. Auflage. München: Hanser.
- Few, S. (2006). Information Dashboard Design. The Effective Visual Communication of Data. Sebastopol: O'Reilly.
- Findlater, L., McGrenere, J., & Modjeska, D. (2008). Evaluation of a role-based approach for customizing a complex development environment. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '08). ACM, New York, NY, USA, 1267-1270. DOI=10.1145/1357054.1357251.
- Gundelsweiler, F. & Reiterer, H. (2008). *Advanced User Interfaces for Product Management Systems*. Proceedings of the 3rd IASTED International Conference on Human Computer Interaction (IASTED-HCI 08), Acta Press, Canada, p. 180-188, Jun 2008.
- Gupta, H. (2011). *Management Information System (An insight)*. New Delhi: International Book House.
- Mahatody, T., Sagar, M., & Kolski, C. (2010). State of the Art on the Cognitive Walkthrough method, its variants and evolutions. *Intl. Journal of Human-Computer Interaction*, 26(8), 741-785.
- Merminod, V., & Blanco, E. (2008). PLM opportunities and weaknesses to support collaborative engineering. In: *proceedings of International Product Development Management Conference*.
- Philpotts, M. (1996). An introduction to the concepts, benefits and terminology of product data management. *Industrial Management & Data Systems* 96 (4): 11-17.
- Rasmussen, N., Chen, C.Y., Bansal, M. (2009): *Business Dashboards, A Visual Catalogue for Design and Development*, Hoboken, NJ: John Wiley & Sons.
- Ristic, S., Rakic-Skokovic, M., Ali, A. (2011) An Overview of the Approches for A PLM Application's Customization. *International Scientific Conference on Industrial Systems*, Serbien 2011

- Rosson, M.B. & Carroll (2002). J.M. Usability Engineering – Scenario-based development of human-computer interaction. San Francisco: Morgan Kaufmann.
- Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & cognition*, 17(6), 759-769.
- Sears, A. (1997). Heuristic Walkthroughs: Finding the Problems Without the Noise, *International Journal of Human-Computer Interaction*, 9, 3, S. 213-234.
- Segawa, S., Sugimura, M., & Ishigaki, K. (2005). New Web-usability evaluation method: Scenario-Based Walkthrough. *Fujitsu scientific and technical journal*, 41(1), 105-114.
- Sendler, U. (2009). Das PLM-Kompendium. Referenzbuch des Produkt-Lebenszyklus-Managements. Heidelberg: Springer.
- Shamsuzzoha, A., Ferreira, F., Azevedo, A., Faria, J., & Helo, P. (2013). Business Process Monitoring and Management in Virtual Enterprise Through Interactive User Interface Layer. In *Advances in Sustainable and Competitive Manufacturing Systems* (pp. 451-464). Springer International Publishing.
- Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*. Washington: IEEE Computer Society Press.
- Stark, J. (2005). Product Lifecycle Management – 21st Century Paradigm for Product Realisation. London: Springer.
- Tille, R., & Burmester, M. (2013 in Druck). Rollenbasierte Interaktion – Entwicklung von Interfacekonzepten für das Product-Lifecycle-Management. *Im Tagungsband der Mensch und Computer 2013 in Bremen*.
- Velcu-Laitinen, O., Yigitbasioglu, O.M. (2012): The Use of Dashboards in Performance Management: Evidence from Sales Managers. *The International Journal of Digital Accounting Research*, vol. 12, 39-58.

Approach to support the modeling of workflows by means of unmoderated data collection sessions

Daniel Ley¹ and Alexander Nielen²

Keywords: data collection, expert modelers, novice modelers, workflow modeling, Subject Matter Experts (SME)

Abstract

A critical component in the design, evaluation and anticipation of human-machine systems is the analysis of Subject Matter Experts' (SME) workflows. Several standard methods, e.g., workshops, are often used to collect corresponding workflow data (Kirwan & Ainsworth, 1992; Stanton et al., 2005). The Process Interviewer (PI) system has been developed to support the modeling of workflows by SMEs by means of unmoderated data collection sessions (Ley, 2011a). Among other things, the system automatically creates feedback for SMEs in terms of graphical workflow models. This can help SMEs (novice modelers) to get an understanding of process modeling as well as analysts (expert modelers) to get an understanding of the corresponding work domain. In the first part of this paper, the approach of the PI system is described. In the second part, the extent to support SMEs to externalize workflow data is highlighted based on fundamental modeling operations identified by Nielen et al. (2012).

Workflow data gathering by means of the PI system

In a first step of system development, information requirements were identified based on common process modeling languages: Business Process Modeling and Notation (BPMN; OMG, 2011), Event-driven Process Chains (EPC; Scheer, 1999) as well as Coordination, Cooperation and Communication (C3; Killich et al., 1999). As a result, a data structure has been developed, distinguishing classes of data relevant for the structure of processes (e.g. activities and process patterns, e.g. XORs) and classes of data relevant for the behavior of processes (e.g. activity duration). Related data requirements were clustered content-wise and finally led to 14 interview steps (Figure 1).

The basic approach of data collection is based on the Hierarchical Task Analysis (Annett, 2004) and uses the structure of so-called "Single Entry Single Exit" (SESE) fragments (Polyvyanyy et al., 2009; described concerning the PI system in Ley, 2011b). Representing a process by means of a process structure tree, four levels can be distinguished: The highest level states the system in which workflows take place. On the level below, general tasks are defined. From there, tasks are iteratively split into activities and sub-activities up to atomic activities.

¹ Fraunhofer FKIE, Germany

² Institute of Industrial Engineering and Ergonomics at RWTH Aachen, Germany

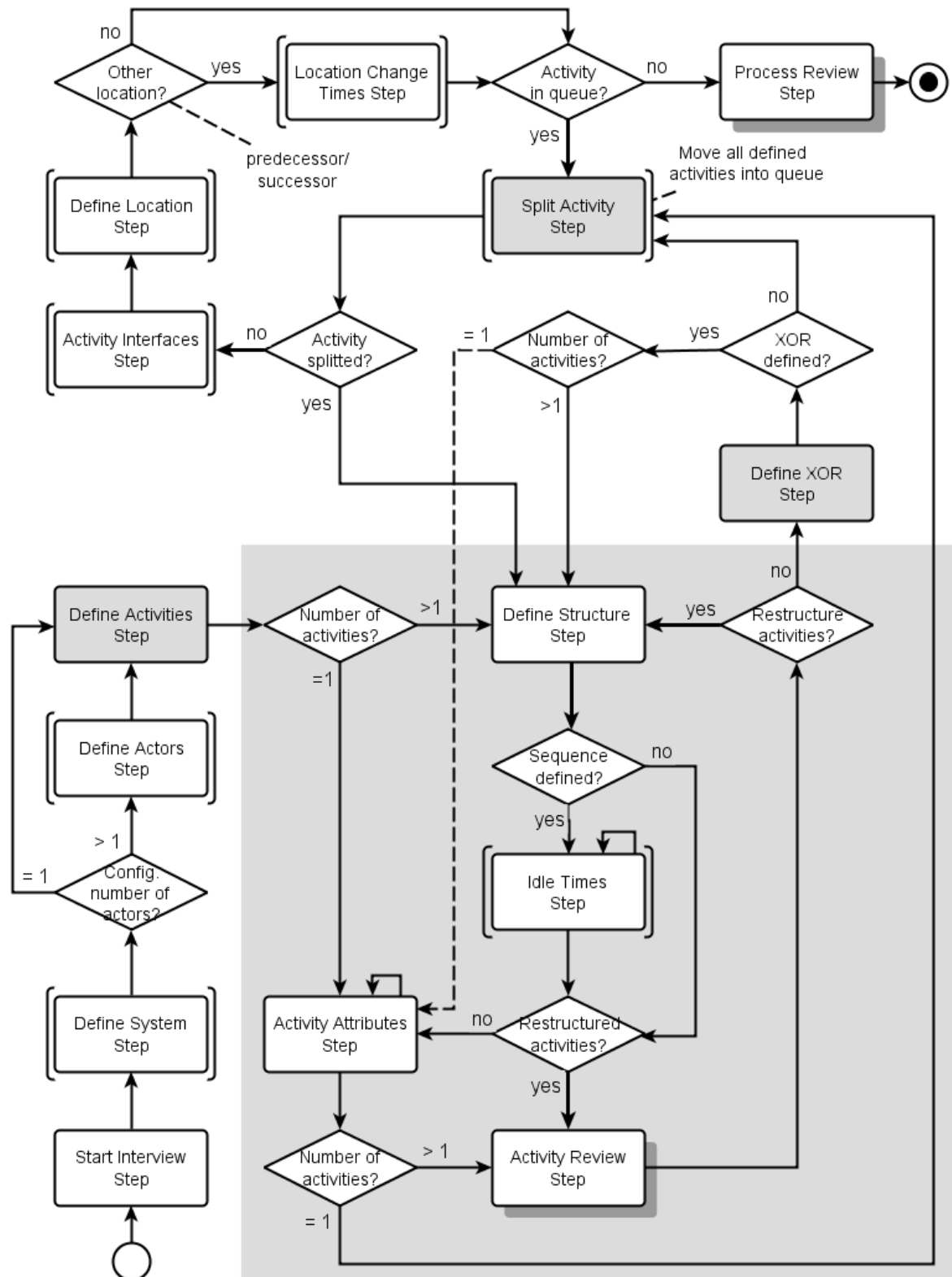


Figure 1: Hybrid program flow of the Process Interviewer (PI) system (basic procedure and steps in which activities are defined are highlighted in grey; steps with graphical process user feedback are shaded)

Atomic activities represent process steps which cannot be further subdivided, while activities above represent (partial) processes. Data capturing takes place top-down and proceeds depth-first.

The chronological order of steps during an interview partly depends on the SME's input and partly is specified by the system based on its configuration. Due to specified and conditional parts, the procedure is called 'hybrid'. The procedure (Figure 1) has been iteratively developed by means of several expert evaluations. SMEs do not need to be aware of the procedure as a whole (dialog principle “controllability” in ISO 9241-110:2006; see Schlick et al., 2010). Basically, there are optional (steps illustrated with brackets) and required steps. The analyst configures the system dependent on the purpose of the analysis and by this means decides which steps are part of the interview. Additionally, the analyst also decides which data are inquired (optional or required) by the system within each interview step.

The procedure consists of three different superior stages. Beginning and end of a phase are determined by steps in which SMEs define new activities or at least are asked to do so (grey shaded steps ‘Define Activities Step’, ‘Define XOR Step’ and ‘Split Activity Step’). By these steps, new partial tree structures are created. In the beginning phase basic information is gathered, like user data (Start Interview Step), the considered system (Define System Step; if not already defined by the analyst) and actors involved in the process (Define Actors Step). Furthermore, tasks are defined (Define Activities Step). In the main phase, (area shaded in grey) if there are more than one defined activity, they are structured (Define Structure Step) as a partial process. In the ‘Idle Times Step’ delays are inquired within activity sequences. After that, activity attributes (e.g. activity time) are captured in the Activity Attributes Step. Now, the partial graphical process is generated and presented by the system in the Activity Review Step, so the SME can interactively revise it. Furthermore, a restructuring of the partial process is possible, which leads into the Define Structure Step again. Otherwise, the SME is explicitly asked for the definition of XOR activities regarding one or several activities of the just entered partial process. In this case, the main phase is launched again; otherwise the first activity of the actual partial process can be splitted into sub-activities (Split Activity Step). The other activities are stored in a queue and can be splitted later on. If an activity is splitted, the main phase for the sub-activities is activated again. Otherwise the atomic phase is launched. In the latter case, interfaces (input, output and means within the Activity Interfaces Step) as well as locations (within the Define Location Step) and possibly the duration of a location change (within the Location Change Time Step) are captured. When finally all activities are passed through the atomic phase (i.e., no more activities are left in the queue of the Split Activity Step) the SME gets feedback in the form of a model of the whole process. After that, the procedure terminates.

Supporting the modeling of workflows by the PI system

The support of process modeling by means of the PI system is now revealed based on nine fundamental modeling operations identified by Nielen et al. (2012; Figure 2). For each operation corresponding actions of the software-based modeling approach (SM) by means of the PI system are outlined and compared to a manual modeling approach (MM) via shapes, marker and whiteboard (Nielen et al., 2012).

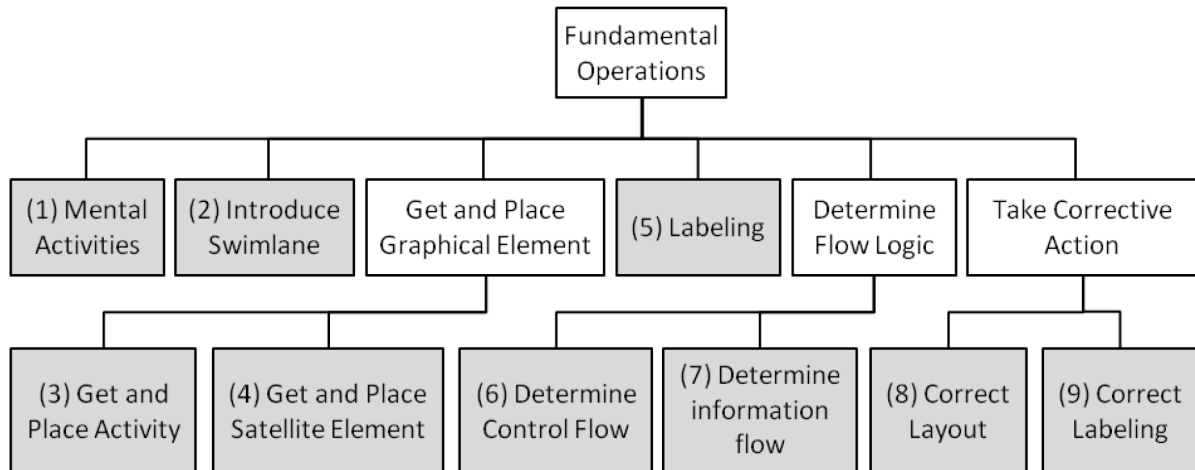


Figure 2: Fundamental operations of process modeling (Nielen et al., 2012)

(1) Mental Activities: Mental (i.e. non-physical) activities concern information processing as well as tasks of proving, deciding and planning (Nielen et al., 2012). A comparison of mental activities between the MM and the SM approach can only be hypothetical. Nevertheless, it can be stated, that mental activities during the MM approach concern the whole process; no assistance is available (unless an analyst takes part). The SME is confronted with the entire process, its elements and unrestricted possibilities of modeling operations. Therefore, SMEs might be probably highly demanded. In contrast, mental activities in the SM approach are focused on single questions in appropriate interview steps, which are less complex. Besides, the hierarchical structuring and successive capturing of workflows allows a focused thinking of the SME concerning one specific part of the workflow. Finally, usable interfaces based on considered and evaluated numerous ergonomic criteria should lead to a certain support of mental activities.

(2) Introduce Swimlane: The purpose of modeling swimlanes is the allocation of activities to specific actors. To model swimlanes via the MM approach, the SME has to know how to sketch swimlanes and must decide whether a horizontal or a vertical representation is pursued. In addition, the arrangement of actors to different swimlanes must be determined. For this, the SME requires an overview of all actors included to the process right from the beginning, since subsequent amendments of sketched swimlanes (including already modeled partial processes) are highly sophisticated. In the SM approach the allocation of activities to actors is performed by the definition of actors in the Define Actors Step and the allocation of actors to a specific activity in the Activity Attributes Step. The PI system adapts

the order of swimlanes and process representation if actors are subsequently defined or deleted in the Activity Attributes Step. No sketching of swimlanes must be accomplished by the SME. The process representation is always horizontal; hence, space is saved and no decision of the SME is needed. In addition, it is possible to allocate several actors to one activity (joint tasks). Contrary to the MM approach, the joint task activities are duplicated by the system for a corresponding swimlane representation.

(3) Get and Place Activity: Getting and placing activities in the MM approach means that the SME selects labeled activities and places them next to other activities in the right swimlane on the whiteboard. This is the preparation to define a specific control flow in a next step. Thereby, the SME can choose one or several activities out of all activities defined so far. In the SM approach, the user selects only two activities (connected to each other in the workflow) of the actual partial process by clicking/touching them. The system takes the placing of activities in a subsequently emerging overlay window for a process pattern specification (see “Determine Control Flow”). The allocation of the activities to a specific swimlane is accomplished later in the Activity Attributes Step. Thus, the extent of the operations is less complex due to the division of the activity selection from swimlane allocation.

(4) Get and Place Satellite Element: Labeled satellite elements must be placed in the MM condition angular (tools) or next to (information) activities. In the SM approach, tools and information, distinguished in input and output, are requested by the system for a specific activity in the Activity Interfaces Step. Hence, satellite elements are directly assigned to activities and no syntax error can occur. Additionally, defined information and tools are stored by the system as classes and can be reused. In the MM condition, they have to be defined multiple times in corresponding cases.

(5) Labeling: The labeling of activities, information and tools is not supported by the SM approach comparing to the MM approach. However, it should be mentioned that typing labels in the SM approach implies that the analyst is always able to read the labels for a subsequent workflow analysis. Labeling is splitted into activity labeling (Define Activities Step, Define XOR Step and Split Activities Step) and labeling of information and tools (Activity Interfaces Step).

(6) Determine Control Flow: Determining the control flow in the MM approach means to place elements of process patterns, like XORs, and the start and the end of a process on the whiteboard, as well as connect these elements with placed activities by drawing lines in the right direction. To do this, the SME has to know different kinds of process patterns and corresponding representations, like decision and merge, split and join, iteration and synchronous collaboration. In the SM approach, the SME only has to designate the process pattern of a component pair (phase 2 of the Define Structure Step), which was previously selected (phase 1). Thus, the SME can focus on one component relation. Syntactical errors are impos-

sible, since the placing of pattern elements and drawing of control flows is accomplished by the system.

(7) Determine Information Flow: The determination of information flows in the MM approach means to connect information elements with a corresponding activity by drawing a dashed line with an arrow on the correct side. Due to the activity specific definitions of information, the system in the SM approach knows the assignments of information to activities and, therefore, can accomplish the determination of information flows. Here, too, the SME can focus on one activity information and does not need to know how to paint the corresponding flow.

(8) Correct Layout: Correcting the layout in the MM approach may imply comprehensive actions, like replacing activities, pattern and satellite elements, as well as redrawing control and information flows. If amendments regard only a limited part of the process, corrections are less complex. In the SM approach, layout revision is accomplished by a restructuring of the entire current partial process in the Define Structure Step. In the case of comprehensive revision demands, this is a suitable procedure. Indeed, it implies more effort in the case of minor correction needs.

(9) Correct Labeling: According to the operation of labeling, the SM approach does not support the revision of labels as well. Though, the possibility to correct labels is given in several interview steps.

(1) to (9): Overall advantage of the SM approach compared to the MM approach is that the SME is guided through the modeling process. Syntactical errors are impossible. Questions are simpler and often focused on a single activity, required actions are restricted, and the system accomplishes many of the structuring and sketching tasks in the process of modeling. Because of this, it is supposed, that a model developed by the SM approach contains much less errors than a model developed by the MM approach. Presuppositions for this hypothesis are that there is no help of analysts and the SME is a novice modeler. Principally, general computer literacy is required.

Summary and future work

In this paper, the interview procedure of the PI system has been outlined. Nielen et al. (2012) identified nine fundamental operations in the process of workflow modeling. The effort of each operation in the manual approach has been compared to the software-based approach with the PI system. Due to various assistance features fewer errors in models created with the PI system are supposed. Similarly to the error classification of Nielen et al. (2011), an empirical study is planned, in which the manual approach is compared to the software-based approach concerning the dependent variables ‘model development time’, ‘activity labeling quality’, ‘omission errors’, ‘execution errors’ and ‘subjective workload’.

Among other things, these are general advantages of the PI system approach: System outcomes are process models, so the analyst does not need to transform data into graphical flowcharts. Besides, these flowcharts provide process knowledge for the analyst. This should provide a basis for the further cooperation of SME and analyst, e.g., in validation sessions. The PI system is able to gather workflow data from various domains. It can be dispatched to distant application places (which saves travel expenses) and allows autonomous and anonymous data captures.

Literature

- Annett, J. (2004). Hierarchical Task Analysis. In: D. Diaper and N. Stanton: *The Handbook of Task Analysis for Human-Computer Interaction*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 67-82.
- ISO 9241-110:2006: Ergonomics of human-system interaction – Part 110: Dialogue principles. ISO, 2006.
- Killich, S., Luczak, H., Schlick, C., Weissenbach, M., Wiedenmaier, S., & Ziegler, J. (1999). Task modelling for cooperative work. *Behaviour and Information Technology*, 18(5), 328-335.
- Kirwan, B., & Ainsworth, L. K. (1992). *A Guide to Task Analysis*. London, Bristol: Taylor & Francis.
- Ley, D. (2011). Approximating Process Knowledge and Process Thinking. Acquiring Workflow Data by Domain Experts. *IEEE Transactions on Systems, Man, and Cybernetics*, October 09-12, 2011, Anchorage, Alaska, USA.
- Ley, D. (2011). Expert-sided Workflow Data Acquisition by means of an Interactive Interview System. *HCD'11 Proceedings of the 2nd international Conference on Human Centered Design held as part of HCI International 2011*, Orlando, Florida, USA. Springer: Berlin.
- Nielen, A., Költer, D., Mütze-Niewöhner, S., & Schlick, C. (2011): Identification and Classification of Human Error in Process Model Development. In: *IEEM 2011, The IEEE International Conference on Industrial Engineering and Engineering Management*, 6 to 9 December 2011, Singapore, 1633-1637.
- Nielen, A.; Mütze-Niewöhner, S., & Schlick, C. (2012). An Empirical Analysis of the Influence of Application Experience on Working Methods of Process Modelers. WASET; ICSSME 2012: *International Conference on Service Science, Management and Engineering*. Eds.: World Academy of Science, Management and Engineering (WASET), Dubai 2012.
- OMG (2011). Business Process Modeling Notation (BPMN). Specification, version 2.0, 2011-01-03.
- Polyvyanyy, A., Smirnov, S., & Weske, M. (2009). On Application of Structural Decomposition for Process Model Abstraction. *Proceedings of the 2nd Inter-*

national Conference on Business Process and Services Computing, Leipzig, Germany, March 2009.

Scheer (1999): *ARIS - Business Process Modeling*, Springer, Berlin.

Schlick, C. M., Bruder, R., & Luczak, H. (2010). *Arbeitswissenschaft*. Berlin, Springer.

Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C., & Jenkins, D. P. (2005). *Human Factors Methods. A Practical Guide for Engineering and Design*. Burlington, USA: Ashgate.

Virtuelle Realitäten in der Psychotherapeutischen Behandlung

Carolyn Wienrich¹, Johann Habakuk Israel² und Gisela Müller-Plath¹

Schlüsselwörter: Mensch-Maschine-Interaktion, Virtuelle Realität, Emotionen, Psychotherapie

Zusammenfassung

Hintergrund. Um einerseits die Diskrepanz zwischen dem Behandlungsbedarf und dem Behandlungsangebot zu minimieren und andererseits die Kosten für psychotherapeutische Behandlungen zu senken, wird der Einsatz virtuellen Realitäten in der Psychotherapie diskutiert. In der hier vorliegenden Arbeit werden Teilergebnisse eines Experimentes vorgestellt, welches das Emotionserleben in einer klinisch-relevanten virtuellen Umgebung untersuchte. Methode. 20 Probanden besuchten ein virtuelles Café, in einer virtuellen Höhe von 0 Meter und 100 Meter. Das Emotionserleben wurde subjektiv, objektiv und physiologisch erfasst. Ergebnisse. Die virtuelle Höhe induzierte spezifische emotionale Reaktionen (Höhenangst) auf allen drei untersuchten Ebenen. Neben Angst vor der Höhe wurde aber ebenso eine unspezifische Aufregung berichtet. Diskussion. Die Bemühung, die Spezifität der emotionalen Reaktionen abzugreifen und von allgemeiner Ängstlichkeit und Aufgeregtheit zu trennen, erwies sich hier als sehr fruchtbar. Dies erscheint insbesondere vor dem Ziel relevant, virtuelle Umgebungen therapeutisch in klinischen Stichproben einzusetzen.

Einleitung

Einem Bericht der Bundestherapeutenkammer (BPtK) zu Folge wird die Zahl der behandlungsbedürftigen psychischen Störungen in Deutschland jährlich auf fünf bis acht Millionen Menschen geschätzt (BPtK, 2011; Bundesministerium für Gesundheit, 2012). Zwischen zehn und zwölf Prozent der Fehltage bei den aktiv Berufstätigen gehen auf Erkrankungen der Psyche zurück (Aktionsbündnis Seelische Gesundheit, 2012; BPtK, 2011). Die Gesamtwirtschaftliche Belastung wird auf 100 Milliarden Euro geschätzt. Diesen Zahlen steht ein Mangel an therapeutischen Behandlungsplätzen gegenüber (Großstadt/Umland: 40/4 Psychotherapeuten für 100.000 Einwohner). Um einerseits die Diskrepanz zwischen dem Behandlungsbedarf und dem Behandlungsangebot zu minimieren und andererseits die Kosten für psychotherapeutische Behandlungen zu senken, wird der Einsatz virtuellen Realitäten in der Psychotherapie diskutiert (ausführliche Diskussion siehe Eichenberg, 2011). Die empirischen Ergebnisse einiger Pilotstudien fielen hinsichtlich der

¹ Technische Universität Berlin, Institut für Psychologie und Arbeitswissenschaften
² Fraunhofer-Institut für Produktionsanlagen- und Konstruktionstechnik (IPK)

Wirksamkeit sehr vielversprechend aus (Metaanalysen und Überblicksartikel: Parson & Rizzo, 2008; Powers & Emmelkamp; Eichenberg, 2011). Indes fehlt eine umfassende Überprüfung, inwieweit virtuelle Realitäten in den Klinikalltag von Patienten und Therapeuten zu integrieren sind. Aspekte der Wahrnehmung, der Bewegungssteuerung, der *Usability* sowie der Produktgestaltung traten bisher nur hintergründig auf. Ziel unserer interdisziplinären Arbeitsgruppe ist es, virtuelle Anwendungen im Bereich der Psychotherapie unter Berücksichtigung dieser Aspekte für einen flächendeckenden Einsatz zu entwickeln, zu evaluieren und diese zu etablieren. In der hier vorliegenden Arbeit wurde untersucht, inwieweit die virtuelle Umgebung *CAVE* in der Lage ist, konkrete emotionale Reaktionen auszulösen und dieser von allgemeiner Aufgeregtheit zu kontrastieren. Als konkreter emotionaler Gegenstand wurde die Höhenangst gewählt. Aufgrund der bisherigen positiven Befunde nehmen wir an, dass auch unsere virtuell dargestellte Höhe, Angst auszulösen vermag. In Ergänzung zu früheren Arbeiten vermuten wir weiter, dass sich diese Angst konkret auf die Höhe zurückführen lässt und kein Artefakt einer unspezifisch gearteten Erregung ist, welche beispielsweise durch das virtuelle Erleben per se entstehen kann.

Methode

An dem Versuch nahmen 20 Probanden (sieben Männer; 19 bis 33 Jahre) teil. Alle Probanden besaßen die Fähigkeit stereoskopisch zu sehen und waren nach Selbstauskunft nicht wegen einer psychischen Störung in Behandlung.

In der *CAVE* des IPK Berlin (www.ipk.fraunhofer.de) wurde ein Szenarium für die Darbietung von Höhe realisiert. Als Grundlage für das virtuelle Modell diente der Kölner *Triangle-Turm* und die 360°-Aussicht von diesem. Präziser ausgedrückt wurde ein Turmcafé modelliert, welches sich im Untergeschoss (*0 Meter (m) Höhe*) oder auf dem Dach (*100m Höhe*) des Turms befinden konnte. In dem Café wurden unter anderen Café-typischen Einrichtungsgegenständen, Pflanzen (*neutrale Position*) mit der Blickrichtung zu einer Wand und Stehtische (*emotionale Position*) mit der Sichtrichtung zu einem großen Fenster dargeboten. Damit wurden die *Höhe*, die *Distanz* zu dem virtuellen Abgrund und die *Blickposition* manipuliert. Außerdem wurde die *Reihenfolge* kontrolliert. Die Hälfte der Probanden besuchten zuerst das Café auf dem Dach, die andere Hälfte das Café im Untergeschoss.

Das Emotionserleben wurde *subjektiv* (Fragebögen), *objektiv* (Beobachtung) und *physiologisch* (Herzfrequenz) erfasst. Für die subjektive Erfassung der Angst als Eigenschaft bzw. als Zustand der Probanden wurde das Trait-State-Angstinventar in der deutschen Fassung von Laux et. al., 1981 verwendet. Zur Erfassung der Höhenangst wurden die folgenden fünf Fragen gestellt, welche in einer Seminararbeit, geleitet von der Erstautorin, entstanden sind: *ha01: ... meide ich hohe Gebäude.; ha02: ... fühle ich mich unwohl, wenn ich in große Tiefen schaue.; ha03: ... habe ich ein flaes Gefühl, wenn ich aus dem Fenster eines hohen Gebäudes schaue.; ha04: ... habe ich Angst vor der Höhe, wenn ich aus dem Fenster eines*

hohen Gebäudes schaue.; *ha05: denke ich, ich könnte sterben, wenn ich aus dem Fenster, eines hohen Gebäudes schaue.*.. Die Antwortmöglichkeiten aller Skalen reichten von 1 (*gar nicht*) bis 5 (*sehr*). Objektiv wurde das emotionale Erleben der Probanden mit Hilfe einer teilnehmenden Beobachtung erfasst. Der Versuchsleiter beurteilte folgende Verhaltensweisen auf der oben beschriebenen vierstufigen Skala: *bo01: kleine Schritte*; *bo02: Zurückzucken*; *bo03: Vorsichtiges Vorbeugen*; *bo03: Schreckhafter Gesichtsausdruck*; *bo07: Zögerndes Annähern zum Abgrund*. Die Herzfrequenz wurde mit dem Forerunner 205/305 von Garmin (<http://www.garmin.com>) aufgezeichnet.

Nach einer zehnminütigen Kalibrierungsphase für das Herzfrequenzmessgerät füllten die Versuchsteilnehmer einen Fragebogen¹ aus. Danach übte jeder Proband die Bewegungssteuerung in der CAVE mit diversen Navigationsaufgaben ein. Wenn eine reibungslose Steuerung gelang, betraten die Teilnehmer das virtuelle Café und akklimatisierten sich für 20 Sekunden (s) an einer Theke. Danach gingen sie zu einer Pflanze, welche 25m von dem Fenster entfernt stand und schauten sie für fünf Sekunden mit der Blickrichtung zur Wand an. Folgend stellten sie sich an einen ersten Stehtisch (25m) und schauten zum Fenster. Anschließend positionierten sich die Teilnehmer vor der Pflanze, welche nahe beim Fenster stand (Blick zur Wand), um nach erneuten fünf Sekunden an den Stehtisch direkt vor dem Fenster mit Sicht aus dem Fenster zu treten. Nachfolgend konnten sich die Teilnehmer zwei Minuten lang frei in dem virtuellen Raum bewegen. Die Probanden wurden via Pfeile und Anweisungen des Versuchsleiters durch die Szenerie navigiert. Bei jedem Ereignis wurden durch den Versuchsleiter Marker für die physiologische Messung gesetzt. Alle Probanden wurden in einem gläsernen Fahrstuhl in das Dachgeschoss des virtuellen Cafés befördert. Während sich der Proband in dem virtuellen Café aufhielt, wurden Geräusche, aufgenommen in einem realen Café, eingespielt. Insgesamt dauerte ein Versuchsdurchlauf circa 60 Minuten.

Aus diesem Ablauf, welcher von allen Probanden komplett durchlaufen wurde (*within-Design*), resultierten 20 kritische Messzeitpunkte. Abbildung 1 zeigt einen Grundriss des virtuellen Cafés sowie vereinfacht die hier relevanten Messzeitpunkte.

Die via den Fragebogen abgefragten Items wurden wie folgt zu Gesamtpunktzahlen innerhalb eines Messzeitpunktes pro Versuchsperson aufsummiert oder gemittelt. Die Trait- und State-Punktwertbildung erfolgte wie im Manual beschrieben und ergab je einen Gesamtwert (*tages* bzw. *sages*). Der Gesamtwert, welcher als Maß der subjektiven Höheangst (*hages*) für die Auswertung herangezogen wurde, resultierte aus einer Mittelung von *ha03* und *ha04*. Die Beurteilungen aller beobachteten Verhaltensweisen (*bo01-bo07*) wurden gemittelt (*boges*). Die Herzfrequenzdaten wurden in Abschnitte von zehn Sekunden gegliedert und gemittelt (*hfges*).

¹ Die hier dargestellten Ergebnisse stellen einen ausgewählten Ausschnitt aus einer größer angelegten Studie dar und konzentrieren lediglich Ergebnisse, welche sich auf Angstmaße und die *Höhe* beziehen.

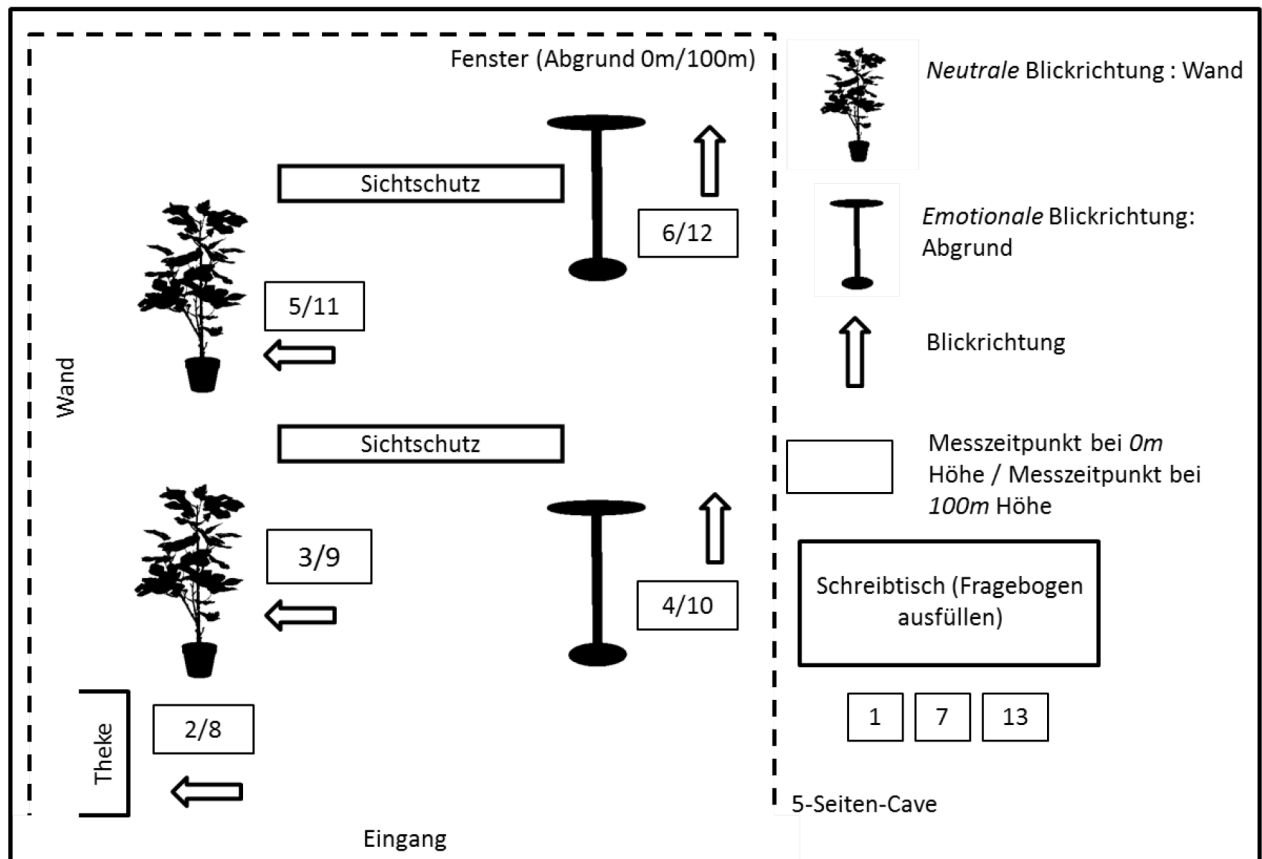


Abb.1: Grundriss des virtuellen Cafés und relevante Messzeitpunkte.

Die Ereignisse konnten mit Hilfe der während des Experimentes gesetzten Marker diesen Mittelwerten zugewiesen werden. Zusätzlich wurden die Herzfrequenzdaten bezüglich der Zeit randomisiert und dann gleich den Originaldaten in zehn Sekunden Sektionen geteilt und gemittelt (*rhfges*). Die randomisierten Daten stellten eine Stichprobe der Herzfrequenzdaten dar, welche sich zufällig auf die Messzeitpunkte verteilen und somit keinen Zusammenhang mit diesen zeigen sollten. Die hier dargestellten Ergebnisse fokussieren den Einfluss der Höhe auf die abhängigen Variablen. Für die statistische Auswertung wurden die folgenden Differenzvariablen zwischen den entsprechenden Messzeitpunkten (abgekürzt dargestellt in eckigen Klammern) für jede Versuchsperson ungeachtet der *Reihenfolge* gebildet:

1. D_SA: sages[13] – sages [7]: allgemeines Angstempfinden,
2. D_HA: hages[12] – hages[6]: konkretes Höhenangstempfinden am Fenster,
3. D_BO: boges[12] – boges[6]: konkretes Angstverhalten am Fenster.
- 4a. Da_HF: hfges[12] – hfges[6]: körperliche Erregung am Fenster,

Die so entstandenen Werte, welche größer als 0 waren, zeigten mehr emotionales Erleben in der Bedingung *100m Höhe* und *0m Distanz* zum Fenster als in der Kon-dition *0m Höhe* und *0m Distanz* zum Abgrund an.

Für die Herzfrequenz bildeten wir weiterhin je eine Differenz pro Höhenbedingung zwischen den Originaldaten und den durch Randomisierung entstandenen Daten:

4b. Db_HF: hfges[12] – rhfges[12],

4c. Dc_HF: hfges[6] – rhfges[6]: körperliche Erregung am Fenster

Um zu testen, ob eine mögliche emotionale Erregung durch den CAVE-Besuch per se ausgelöst und damit unspezifisch hinsichtlich der *Höhe* ist, bildeten wir die entsprechenden Differenzen getrennt für die *Reihenfolge 1: 0m-100m* und *2: 100m-0m*, so dass sich jeweils zwei Differenzwerte pro Variable ergaben: D_SA1, D_SA2; D_HA1, D_HA2; D_BO1, D_BO2; Da_HF1, Da_HF2; Db_HF1; Db_HF2; Dc_HF1, Dc_HF2. Mittels Einstichproben-T-Test prüften wir folgende Hypothesen hinsichtlich des Vergleichs *100m Höhe* versus *0m Höhe* bei einer *Distanz* von *0m* zum Fenster/Abgrund:

H1: $D_SA = \mu_0$, keine Erhöhung der allgemeinen Angstwerte,

H2: $D_HA > \mu_0$, Erhöhung der berichteten Höhenangst bei der *100m* Bedingung,

H3: $D_BO > \mu_0$, Erhöhung der beobachtbaren Höhenangst bei *100m* Bedingung,

H4a: $Da_HF > \mu_0$, **H4b:** $Db_HF > \mu_0$, **H4c:** $Dc_HF = \mu_0$; Erhöhung der physiologischen Erregung bei der *100m* Bedingung, wobei $\mu_0 = 0^1$.

Weiterhin nahmen wir an, dass der Effekt der *Höhe* erhalten bleibt, wenn die *Reihenfolge* Berücksichtigung findet (geprüft mit einem Zweistichproben-T-Test):

H5: $D_HA1 = D_HA2$

H6: $D_BO1 = D_BO2$

H7a: $Da_HF1 = Da_HF2$; **H7b:** $Db_HF1 = Db_HF2$; **H7c:** $Dc_HF1 = Dc_HF2$

Ergebnisse und Diskussion

Die Gesamtwerte des State-Trait-Angstinventars vor dem Besuch des virtuellen Cafés zeigten, dass unsere Stichprobe Trait-Werte im Bereich der Normstichprobe erzielte ($M_{\text{Stichprobe}}=32.75$; $SD_{\text{Stichprobe}}=6.75$ vs. $M_{\text{Norm}}=35.07$; $SD_{\text{Norm}}=9.05$). Es gab zwei Probanden, welche als höher ängstlich eingestuft werden konnten und sieben, welche weniger Ängstlichkeit angaben als die entsprechende Normgruppe. Bezüglich der spezifischen Höhenangstitems zeigten 13 Versuchsteilnehmer *keine* oder *ein wenig* Höhenangst und sieben Probanden *ziemliche* oder *starke* Angst vor der Höhe. Abbildung 2 und Tabelle A1 illustrieren die Ergebnisse der Hypothesentestung. Wie erwartet unterschied sich das allgemeine Angstempfinden (*sages*) nicht zwischen den beiden Höhenbedingungen. Ebenso unserer Erwartung entsprechend, wurde nach der Darbietung des *100m*-Szenariums mehr Angst bei

¹ Kein Unterschied zwischen den beiden Ausprägung des Faktors Höhe bei einer *Distanz* von *0m* zum Fenster.

den spezifischen Höhenangstitems berichtet als nach dem *0m*-Szenarium. Dieses Ergebnis wurde durch die Beobachtungsdaten bestätigt. Auch die physiologischen Daten indizieren mehr Erregung während die *100m*-Bedingung durchlaufen wurde im Vergleich zur *0m*-Kondition. Allerdings fiel hier das Ergebnis nicht so eindeutig aus. Im Vergleich zu den randomisierten Herzraten waren zwar die Originaldaten wie erwartet signifikant höher in der *100m*-Bedingung, aber dieser war tendenziell auch bei *0m* Höhe zu beobachten. Die Differenz zwischen *100m* und *0m* Höhe unterschied sich hingegen nicht signifikant von null und widersprach damit unseren Erwartungen.

Ein genauerer Blick auf die Herzfrequenzdaten ergab, dass insbesondere Versuchspersonen eine höhere physiologische Erregung in der Bedingung zeigten, die sie zuerst besuchten (H7a-H7c). Da sich diese Ergebnisse nicht gleichermaßen signifikant bezüglich der Höhenangst- und Beobachtungsitems feststellen ließen, kann man einerseits vermuten, dass die erhöhten Herzfrequenzen, nicht alleinig Angstempfinden, sondern ebenso eine generelle Erregung widerspiegeln. Mit anderen Worten nicht nur die virtuelle Höhe, sondern auch der Cave-Besuch per se könnte bei den Probanden zu einer physiologischen Reaktionen geführt haben. Werden allerdings die Antworten der offen gestellten Frage: *Was haben Sie empfunden, als Sie direkt am Fenster standen?*, in Betracht gezogen, kann man andererseits zu der Schlussfolgerung kommen, dass es drei Gruppen von Probanden gab. Ein kleiner Teil (vier Probanden) gab keine besonderen Empfindungen an. Neun Teilnehmer empfanden das Erlebnis als *aufregend*, *spannend* und *interessant*. Sechs Probanden gaben hingegen an, dass sie *Angst vor der Höhe* empfunden hätten und ein *mulmiges Gefühl* hatten. Interessanterweise berichteten die Teilnehmer vor dem Versuch mehr Höhenangst, welche am Fenster *Angst vor der Höhe* ($M=3$; $SD=.99$) berichteten als jene, die es als *spannendes Erlebnis* ($M=2$; $SD=1$) empfunden haben ($t(13)=1.98$; $p=.04$).

Dieser knappe Ergebnisbericht lässt einige Frage, darunter eine Einschätzung der Benutzerfreundlichkeit, die Generalisierbarkeit auf andere Emotionen und der Einsatz beim Vorliegen einer klinisch relevanten Höhenangst ungeklärt. Es bleibt also genauer zu untersuchen, welche Faktoren bei der Gestaltung von virtuellen Umgebungen wichtig sind, damit sie in einem klinischen Setting zur Anwendung gebracht werden können. Erkenntnisse aus den Ingenieurwissenschaften, der Psychologie sowie Usability-Forschung müssen hier weiter zusammengebracht werden, um diese besondere Schnittstelle der Mensch-Maschine-Interaktion erfolgreich zu gestalten. Folgestudien der hiesigen interdisziplinären Arbeitsgruppe möchten dazu einen weiteren Beitrag leisten.

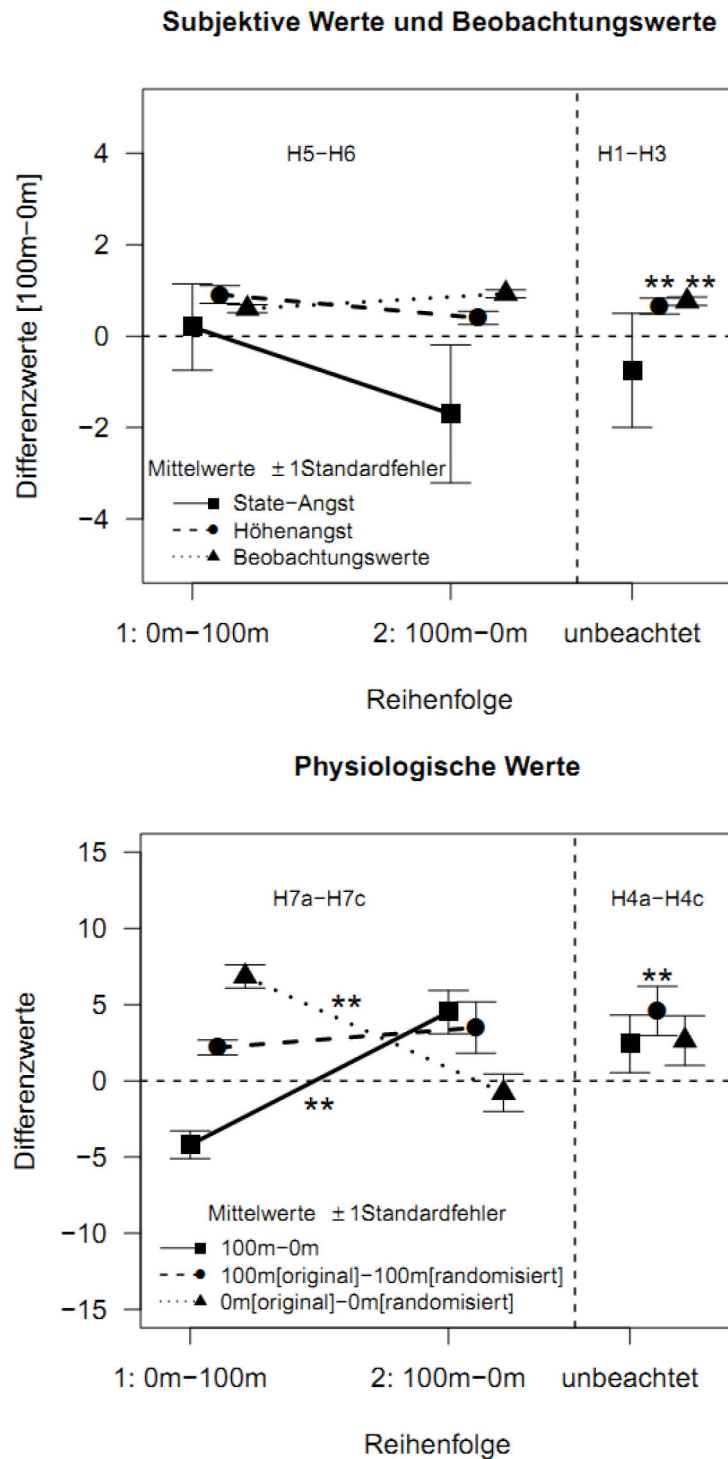


Abb.2: Ergebnisse. Obere Grafik: rechter Abschnitt H1-H3; linker Abschnitt H5-H6. Untere Grafik: rechter Abschnitt H4a-H4c; linker Abschnitt H7a-H7c. **: $p \leq .01$.

Schlussfolgerung

Zusammenfassend konnte festgestellt werden, dass die virtuelle Umgebung Cave spezifische emotionale Reaktionen (Höhenangst) induzieren konnte. Unter den Probanden fanden sich allerdings auch solche, welche die virtuelle Höhe als spannend und aufregend empfanden. Die Bemühung, die Spezifität der emotionalen Reaktionen abzugreifen und von allgemeiner Ängstlichkeit und Aufregtheit zu

trennen, erwies sich hier als sehr fruchtbar und sollte auch in zukünftigen Studien stärker Beachtung finden. Dies erscheint insbesondere vor dem Ziel relevant, virtuelle Umgebungen therapeutisch in klinischen Stichproben einzusetzen.

Literatur

- Aktionsbündnis Seelische Gesundheit. (2012). *Psychische Erkrankungen*. Verfügbar unter <http://www.seelichegesundheit.net>
- BPtK. (2011). Wartezeiten für GKV-Versicherte BPtK fordert höchstens drei Wochen Wartezeit für psychisch kranke Menschen. Verfügbar unter <http://www.bptk.de>
- Eichenberg, C. (2011). *Virtual reality*. In J.-J. Kim (Hrsg.), S. 481-496. InTech.
- Laux, L., Glanzmann, P., Schaffner, P., & Spielberger, C. D. (1981). *Das State-Trait-Angstinventar*. Weinheim: Beltz.
- Parsons, T., & Rizzo, A. A. (2008). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, 39, S. 250-261.
- Powers, B. M., & Emmelkamp, M. G. (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders*, 39, 250-261.

Anhang

Tab.A1: Hypothesen und Ergebnisse der Hypothesentestung.

	M ^a	SD ^b	t (df) ^c	p ^d
H1	-.75	5.71	-.58 (19)	.56
H2	.65	.80	3.65	< 0.01
H3	.76	.42	8.18 (19)	< .01
H4a	2.42	8.68	1.25 (19)	.11
H4b	4.59	7.39	2.78 (19)	< .01
H4c	2.64	7.47	1.59 (19)	.07
	M[0m-100m]; M[100m-0m]	SD[0m-100m]; SD[100m-0m]	t (df)	p
H5	.91; .40	.88; .67	-1.46 (18)	.16
H6	.93; .60	.39; .40	1.87 (18)	.08
H7a	9.03; 4.19	4.15; 6.68	5.31 (18)	< .01
H7b	-6.99; 2.19	2.21; 9.87	1.50 (18)	.08
H7c	-1.56; 6.85	3.51; 8.13	-3.00 (18)	< .01

^aMittelwert; ^bStandardabweichung; ^cT-Wert(Freiheitsgrade); ^dp-Wert

Danksagung

Ein Dank gilt den studentischen Praktikanten Sebastian Runge und Jan Richter für die Modellierung der Szenerien und die technischen Unterstützung während der Versuche.

Beanspruchung bei Head-Mounted Displays im Langzeiteinsatz

Matthias Wille, Britta Grauel und Lars Adolph¹

Schlüsselwörter: Head-Mounted Displays, Datenbrillen, Beanspruchung, IKT

Zusammenfassung

In einer Laborstudie zur langfristigen Beanspruchung beim Einsatz von Head-Mounted Displays (HMDs) arbeiteten 41 Probanden für jeweils 4 Stunden mit einem HMD oder vergleichsweise mit einem Tablet-PC. Ihre Aufgabe war es, in einer bildbasierten Montageanweisung Fahrzeugmodelle aufzubauen und gleichzeitig einer Monitoringaufgabe nachzugehen. In den Ergebnissen zeigte sich, dass mit dem HMD langsamer in der Montageaufgabe gearbeitet wurde und dass bei der Monitoringaufgabe die Trefferquote beim HMD tendenziell geringer ist, während die Reaktionszeiten vergleichbar blieben. Mehrere prozessbegleitend erhobene subjektive Beanspruchungsparameter konnten eine höhere Beanspruchung beim Arbeiten mit dem HMD nachweisen. Ein Fragebogen zur visuellen Ermüdung zeigte ebenfalls einen stärkeren Effekt bei HMDs und einen stärkeren Anstieg über die Zeit. In Interviews wurde das Arbeiten mit dem HMD oft als unangenehm beurteilt und das Arbeiten mit dem Tablet-PC bevorzugt. Die Ergebnisse sind dabei wahrscheinlich teilweise auf die Hardwareausführung des HMDs und nicht nur auf die Technologie als solche zurückzuführen.

Einleitung

Bei Head-Mounted Displays (HMDs) – kleine Monitore, die direkt vor den Augen platziert sind – gibt es verschiedene Aufbautypen und Anwendungsfelder (für einen Überblick siehe z. B. Wille et al. 2011). Im Rahmen des Forschungsprojektes "Head-Mounted Displays – Bedingungen des sicheren und beanspruchungsoptimalen Einsatzes" untersucht die Bundesanstalt für Arbeitsschutz und Arbeitsmedizin den Einsatz von HMDs als adaptives Arbeitsassistenzsystem. Der Vorteil liegt hier vor allem darin, dass der Nutzer bei der Ausführung der Arbeitsaufgabe beide Hände frei hat. Der Fokus der hier berichteten Untersuchungen liegt auf einem länger andauernden Einsatz über mehrere Stunden, da hier noch wenige Erkenntnisse über die Beanspruchungsverläufe vorliegen.

Stichprobe

Insgesamt nahmen 41 Probanden in einem Alter von 18-67 Jahren an der Untersuchung teil. Die Stichprobe wurde mittels Mediansplit in 2 Altersgruppen unterteilt, um so möglichen Alterseffekten nachzugehen. In der Gruppe der Jüngeren waren

¹ Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Gruppe Human Factors, Ergonomie

21 Probanden im Alter zwischen 18-31 Jahren ($Mw = 23,48$; $Sd = 3,341$), wovon 9 männlich und 12 weiblich waren. In der Gruppe der Älteren waren 20 Probanden im Alter von 38-67 Jahren ($Mw = 50,55$; $Sd = 9,214$), wovon 11 männlich und 9 weiblich waren. Keiner der Probanden hatte vorher schon mit einem HMD gearbeitet.

Versuchsaufbau

Um die Beanspruchung beim länger andauernden Einsatz von HMDs zu untersuchen, wurde in den Laborräumen der BAuA eine Studie durchgeführt. Hierbei arbeitete jeder Proband in 3 Sessions jeweils 4 Stunden entweder mit einem HMD oder vergleichsweise mit einem Tablet-PC. Sie hatten dabei zwei Aufgaben zu erfüllen, die beide als gleich wichtig instruiert wurden ("Dual-task" Paradigma). Zum einen sollten sie Modellfahrzeuge aus einzelnen Bausteinen zusammenbauen, wobei die einzelnen Montagefolien der Aufbauanleitung entweder auf dem HMD oder auf dem Tablet-PC angezeigt wurden. Diese bildbasierte, schrittweise Aufbauanleitung wurde gewählt, da sie industriellen Montageanleitungen ähnelt, wie sie z.B. in der Instandhaltung verwendet werden. In einer parallel dargebotenen Monitoring-Aufgabe sollten die Probanden auf 3 vertikale Balken am äußeren Bildschirmrand achten und bestimmte Zustandsänderungen mit Knopfdruck quittieren. Die Balken variierten kontinuierlich, aber sehr langsam in ihrer Länge und ab und zu in ihrer Farbe (blau/rot). Es sollte mit Knopfdruck zum einen ein Farbwechsel bestätigt werden, der durch die großflächige Veränderung in der Darstellung einen visuellen Popout-Effekt beinhaltet. Zum anderen sollte ein Positionswechsel des längsten Balkens, der durch die kontinuierliche Längenvariation der Balken entsteht, mit einem anderen Knopf quittiert werden. Hierbei bekamen die Probanden in der Hälfte der Blöcke eine Rückmeldung über ihre letzte Bestätigung in Form einer schriftlichen Mitteilung über den Balken (z. B. "Mitte - Rot").

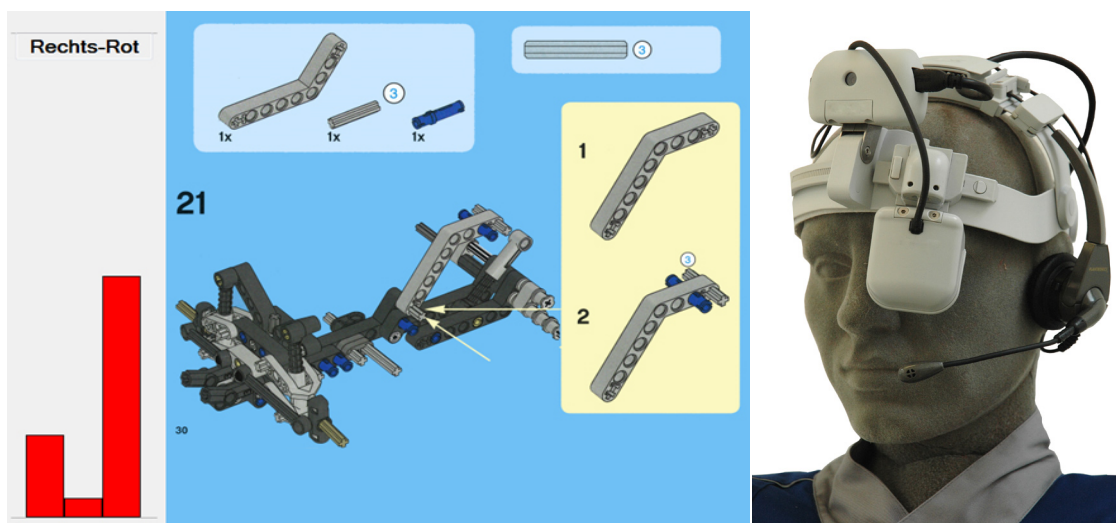


Abb. 1: Links: Darstellung einer Montagefolie und der Balkenaufgaben im HMD (hier für linkes Führungsauge); Rechts: Verwendetes HMD der Firma Heitec.

Das verwendete HMD ist ein MAVUS¹-System der Firma Heitec, wie es derzeit auch in der industriellen Instandhaltung eingesetzt wird. Es ist ein monokulares "look around" Display mit einer Auflösung von 800 x 600 Pixeln. Die Technik ist an einem Kopfträger befestigt, der auch eine Frontkamera und ein Headset beinhaltet, welche allerdings bei dem Versuch keine Funktion erfüllten. Im industriellen Einsatz sind Headset und Kamera für die Kommunikation mit Experten vorgesehen. Der Kopfträger wiegt 380 Gramm und ist über ein Kabel mit einer Weste verbunden, die die Funktechnologie zur Übertragung der Daten und den Akkumulator für die Stromversorgung beinhaltet. Als Tablet-PC kam ein CL900 der Firma Motion² zum Einsatz. Um zu gewährleisten, dass die Darstellung der Arbeitsinhalte gleich ausfällt, wurde auch hier nur ein Fenster von 800 x 600 Pixeln dargestellt und die übrige Fläche abgeklebt. Sämtliche Interaktionen (Vor- und Zurückschalten der Montagefolien, Tasten zur Quittierung der Balkenaufgaben) fanden über ein umgebautes Nummernpad der Firma Logitech statt.

Die Probanden saßen in einem gut ausgeleuchteten Raum vor einem höhenverstellbaren Tisch. Zu Beginn des Versuches wurde der Arbeitsplatz auf die Körpermaße des Probanden eingestellt. Die Versuchsleiter saßen hinter einer Trennwand und überwachten den Versuch, der auch mittels Video komplett aufgezeichnet wurde. Jeder Proband machte 3 dieser Versuchssessions an 3 verschiedenen Tagen. Wobei zwischen den Sessions mindestens 2 und maximal 10 Tage lagen. Die erste Session wurde immer mit dem HMD durchgeführt. Für die 2. und 3. Session wurde die Vergleichsmessung mit dem Tablet-PC und die Wiederholungsmessung mit dem HMD permutiert. Die Wiederholungsmessung mit dem HMD fand statt, um so erste Gewöhnungseffekte an die neue Technologie abzuschätzen.

Versuchsablauf und erhobene Variablen

Am Anfang des Versuches wurde zunächst ein Fragebogen zur Technikaffinität (TA-EG, Karrer et al., 2009) ausgefüllt und die Instruktion in schriftlicher Form gegeben. Dann wurde die RSME-Skala (Zijlstra, 1993) zur mentalen Beanspruchung vorgestellt, die später zu verschiedenen Zeitpunkten im Versuch abgefragt wurde. Anschließend wurde den Probanden das HMD gezeigt und die Funktionen erläutert. Es wurde der Kopfträger an die Probanden angepasst und das HMD eingestellt. Hierbei wurde das monokulare HMD immer vor dem individuellen Führungsauge des Probanden montiert und die Balken immer im äußeren Bereich des Blickfeldes dargestellt. In der ersten Session wurde dann zunächst die Kombination der beiden Aufgaben für ca. 10 Minuten geübt. Der anschließende Versuch erfolgte in 4 Blöcken von jeweils 50 Minuten. Zwischen den ersten und letzten beiden Blöcken war eine kurze Pause von 5 Minuten, in der das HMD aber anbehalten wurde. Zwischen Block 2 und 3 – in der Mitte des Versuches – gab es eine größere

¹ <http://www.mavus.info> [abgerufen August 2013]

² <http://www.motioncomputing.de> [abgerufen August 2013]

Pause von ca. 15 Minuten, in der das HMD abgenommen wurde und die Probanden aufgefordert waren, sich zur Belastungsvarianz kurz zu bewegen.

Prozessbegleitend wurden mehrere subjektive Beanspruchungsparameter erhoben: Der RSME, der auf einer Skala von 0-150 die empfundene Beanspruchung wiedergibt, wurde jeweils in der Mitte eines Blockes und am Ende während der Arbeit abgefragt, so dass insgesamt 8 Messzeitpunkte vorliegen, um eine Entwicklung der Beanspruchung über die Zeit wiederzugeben. Der NASA-TLX (Hart & Staveland, 1988) mit seinen 6 Subskalen wurde nicht direkt in der Arbeitssituation, sondern unmittelbar zu Beginn der großen Pause und nach Ende des Versuches erhoben, so dass hier 2 Messzeitpunkte zur Verfügung stehen. Der Paarvergleich zur individuellen Gewichtung der Skalen wurde ebenfalls im Anschluss an den Versuch durchgeführt. Des Weiteren wurde ein Fragebogen zur visuellen Ermüdung (Visual Fatigue Questionnaire nach Bangor, 2000) zu 5 Zeitpunkten erhoben: Unmittelbar vor Beginn des Versuches, um die individuelle Ausgangslage festzustellen, und jeweils direkt nach Ende eines jeden Blockes. (Als objektiver Beanspruchungsparameter wurde auch die Herzrate während des gesamten Versuches erhoben. Diese Daten werden an anderer Stelle dargestellt.) Im Anschluss an jede Versuchssession wurde mit den Probanden ein Interview durchgeführt.

Versuchsplan

Es liegt ein Messwiederholungsplan mit mehreren abhängigen und unabhängigen Variablen vor. Unabhängige Variablen sind die Versuchssession (1. Messung HMD, 2. Messung HMD, Tablet-PC) und die Rückmeldung in der Balkenaufgabe (blockweiser Wechsel). Das Alter und die Technikaffinität der Probanden sind Zwischensubjektfaktoren (jeweils Mediansplit in 2 Gruppen). Abhängige Variablen sind die Anzahl der bearbeiteten Montagefolien pro Session, die Trefferquote und Reaktionszeit in den Balkenaufgaben, RSME-Werte zu 8 Zeitpunkten, NASA-TLX zu 2 Zeitpunkten, Visual Fatigue Questionnaire zu 5 Zeitpunkten.

Ergebnisse

Ausgewertet wurde in SPSS 20 mittels ANOVA mit Messwiederholung. Hierbei wurden stets 2 Versuchssessions paarweise verglichen. Ein Vergleich über alle 3 Sessions wäre inhaltlich nicht zulässig, da sie nicht Stufen eines Faktors darstellen, sondern es sich im einen Fall um eine Wiederholungsmessung zur Ableitung von Gewöhnungseffekten und im anderen Fall um eine Vergleichsmessung unterschiedlicher Technologie handelt. Aus Gründen der Übersicht wird in diesem Artikel vorrangig auf den Vergleich zwischen 2. HMD-Messung und Tablet-PC eingegangen, der Unterschiede zwischen den Displaytechnologien wiedergibt. Hierbei wurde die 2. Messung mit dem HMD gewählt, da hier bereits genügend Einarbeitung mit der neuen, für die Probanden ungewohnten, Technologie vorhanden ist. Der Vergleich zwischen 1. und 2. Messung mit dem HMD, der die Gewöhnung an die Technologie wiedergibt, wird bei signifikanten Effekten ebenfalls beschrieben.

Als Leistungsindikator in der Montageaufgabe wurde die Anzahl der pro Session abgearbeiteten Aufbauschritte verwendet. Dies ist möglich, da sich die Komplexität einzelner Arbeitsschritte ähnelt und kompliziertere Modelle vor allem aus mehr Arbeitsschritten bestehen. Bei einem Vergleich über beide Altersgruppen hinweg zeigt sich, dass mit einem HMD signifikant weniger Schritte erledigt werden als bei Benutzung eines Tablet-PCs (HMD 1. Messung 130,9; HMD 2. Messung 128,8; Tablet-PC 157,7). Unter Berücksichtigung des Faktors Alter zeigt sich neben dem weiterhin bestehenden Haupteffekt Display [$F(1,37) = 24.285$; $p = .000$] auch ein hochsignifikanter Haupteffekt Alter [$F(1,37) = 23.647$; $p = .000$]. Hierbei gibt der Haupteffekt wieder, dass Ältere weniger Montagefolien abarbeiten (siehe Tabelle 1). Ein Gewöhnungseffekt an das HMD durch den Vergleich zwischen 1. und 2. Messung kann nicht nachgewiesen werden [$F(1,37) = .208$; $p = .651$].

Tab.1: Mittlere Anzahl bearbeiteter Montagefolien nach Versuchssession und Alter (Standardabweichung in Klammern).

	1. Messung HMD	2. Messung HMD	Tablet-PC
Jüngere Probanden	155 (40,12)	165 (42,34)	198 (50,51)
Ältere Probanden	108 (41,92)	94 (46,37)	120 (71,20)

Bei der Monitoringaufgabe zeigt sich ein hochsignifikanter Unterschied zwischen den Aufgaben Balkenfarbe und Balkenlänge: Der Wechsel der Balkenfarbe hat stets eine höhere Trefferquote [$F(1,36) = 84.690$; $p = .000$] und geringere Reaktionszeit [$F(1,36) = 17.560$; $p = .000$] als der Wechsel der Position des längsten Balkens (siehe Tabellen 2 und 3). Dieser Effekt illustriert die Wirkung eines visuellen Popouts durch die großflächige Veränderung beim Farbwechsel und war so auch zu erwarten. Ebenso führt eine Rückmeldung über die letzte bestätigte Position ("FB") zu einem hochsignifikanten Anstieg der Trefferquote [$F(1,37) = 58.257$; $p = .000$] in beiden Balkenaufgaben. Bei der Reaktionszeit zeigt sich aber kein Haupteffekt der Rückmeldung [$F(1,37) = .078$; $p = .782$]. Der Displaytypus zeigt keinen signifikanten Einfluss auf die Trefferquote [$F(1,37) = 3,635$; $p = .065$] oder die Reaktionszeit [$F(1,37) = 2,858$; $p = .100$]. Die Tendenz geht aber in Richtung besserer Werte für den Tablet-PC.

Tab.2: Mittlere Trefferquote in Prozent in den Monitoringaufgaben nach Versuchssession und Rückmeldung (Standardabweichung in Klammern).

		1. Messung HMD	2. Messung HMD	Tablet-PC
Balkenfarbe	ohne FB	78,17 (15,91)	81,02 (14,99)	80,86 (15,12)
	mit FB	87,30 (15,27)	89,85 (13,25)	92,62 (11,18)
Balkenlänge	ohne FB	63,25 (23,43)	66,11 (21,96)	71,39 (18,45)
	mit FB	73,99 (21,58)	79,75 (18,07)	81,23 (14,07)

Tab.3: Mittlere Reaktionszeit in Sekunden in den Monitoringaufgaben nach Versuchssession und Rückmeldung (Standardabweichung in Klammern).

		1. Messung HMD	2. Messung HMD	Tablet-PC
Balken- farbe	ohne FB	9,60 (5,91)	11,50 (11,21)	11,86 (9,47)
	mit FB	13,68 (8,26)	12,85 (8,98)	11,50 (6,65)
Balken- länge	ohne FB	20,29 (21,64)	20,93 (21,36)	17,66 (19,21)
	mit FB	19,43 (12,48)	19,99 (14,17)	16,14 (10,61)

Ein Gewöhnungseffekt an das HMD durch den Vergleich zwischen 1. und 2. Messung zeigt sich hinsichtlich der Trefferquote [$F(1,37) = 5.395$; $p = .026$], aber nicht bezogen auf die Reaktionszeit [$F(1,37) = .196$; $p = .661$]. Ein Effekt des Alters ließ sich im Rahmen der Monitoringaufgabe allgemein nicht feststellen.

Bei der subjektiven Beanspruchung im Rahmen des Gesamtwertes des NASA-TLX (siehe Tabelle 4) zeigt sich ein signifikant höherer Wert für das HMD verglichen mit dem Tablet-PC [$F(1,37) = .26.952$; $p = .000$]. Die Steigerung der empfundenen Beanspruchung über die Zeit wird ebenso signifikant [$F(1,37) = 14.267$; $p = .001$], allerdings liegt keine Wechselwirkung zwischen Displaytyp und Zeitpunkt vor [$F(1,37) = 1.375$; $p = .253$]. Eine Wechselwirkung zwischen Displaytyp und Alter erreicht statistische Tendenz [$F(1,37) = 3.172$; $p = .083$]. Und erstmals zeigt sich auch ein Einfluss von Technikaffinität in Wechselwirkung mit dem Displaytyp [$F(1,37) = 5.000$; $p = .031$] in der Form, dass nicht technikaffine Probanden nicht nur allgemein einen höheren Beanspruchungsscore haben (HMD: 73.18; Tablet-PC: 66.17) als technikaffine Nutzer (HMD: 68.96; Tablet-PC: 51.35), sondern auch weniger von der Entlastung durch den Tablet-PC profitieren. Ein Gewöhnungseffekt in der Beanspruchung zwischen 1. und 2. Messung mit dem HMD lässt sich nicht nachweisen.

Tab.4: Mittlerer Gesamtwert des NASA-TLX nach Versuchssession und Zeitpunkt und Alter (Standardabweichung in Klammern).

		1.Messung HMD	2. Messung HMD	Tablet-PC
Nach 2 Stunden	Jüngere	74,65 (21,02)	67,95 (25,05)	54,20 (25,39)
	Ältere	71,98 (18,66)	66,64 (21,15)	58,69 (22,91)
Nach 4 Stunden	Jüngere	72,83 (21,33)	73,18 (23,83)	55,05 (22,11)
	Ältere	81,14 (26,72)	75,95 (23,69)	65,76 (22,81)

Beim RSME-Wert (siehe Tabellen 5 und 6) zeigt sich ebenfalls ein hochsignifikanter Anstieg der Beanspruchung beim HMD im Vergleich zum Tablet-PC [$F(1,35) = 84.435$; $p = .000$], ein Anstieg über die Zeit [$F(1,29) = 6.956$; $p = .000$] und ein Haupteffekt des Alters [$F(1,35) = 6.684$; $p = .014$], der wiedergibt, dass Ältere beanspruchter sind.

Tab.5: Mittlerer Gesamtwert des RSME für jüngere Probanden nach Versuchssession und Zeitpunkt in Minuten (Standardabweichung in Klammern).

	1.Messung HMD	2. Messung HMD	Tablet-PC
30	46,65 (17,02)	35,45 (15,33)	25,95 (11,87)
60	58,25 (19,08)	44,95 (17,37)	32,00 (13,50)
90	56,00 (23,25)	45,55 (18,94)	31,45 (16,57)
120	56,55 (11,82)	47,50 (17,24)	33,40 (16,15)
150	46,00 (16,41)	39,60 (17,15)	30,95 (13,24)
180	44,55 (12,20)	46,65 (16,77)	33,85 (16,08)
210	49,95 (17,17)	48,35 (18,86)	33,75 (17,84)
240	53,60 (16,29)	50,89 (19,55)	36,05 (18,63)

Tab.6: Mittlerer Gesamtwert des RSME für ältere Probanden nach Versuchssession und Zeitpunkt in Minuten (Standardabweichung in Klammern).

	1.Messung HMD	2. Messung HMD	Tablet-PC
30	51,90 (24,62)	49,48 (18,30)	32,48 (16,52)
60	61,62 (26,96)	55,48 (21,25)	44,00 (22,03)
90	55,43 (20,64)	62,00 (21,53)	43,33 (22,10)
120	65,14 (25,12)	62,71 (21,81)	44,52 (23,60)
150	64,05 (28,92)	60,43 (26,71)	46,00 (28,22)
180	67,50 (31,09)	63,86 (31,69)	49,90 (28,69)
210	70,95 (32,96)	68,57 (30,40)	49,48 (32,88)
240	76,67 (35,99)	73,81 (37,08)	51,62 (31,92)

Der Visual Fatigue Questionnaire (VFQ) erhebt mit 16 Items verschiedene Aspekte visueller Beanspruchung und Ermüdung. Jedes Item ist auf einer zehnstufigen Skala, ein zusammenfassender Gesamtwert existiert nicht. Der Fragebogen wurde vor Beginn des Versuches und dann stündlich erhoben. Es zeigte sich eine deutliche, hochsignifikante Veränderung fast aller Itemscores, die allerdings auf einem geringen Mittelwertsniveau von 1 - 3,5 bleiben. Alle Items zeigen eine Zunahme über die Zeit und eine stärkere Beanspruchung beim HMD. Bei vielen Items wird auch die Wechselwirkung Zeitpunkt*Display signifikant, die einen stärkeren Anstieg der visuellen Beanspruchung beim HMD über die Zeit widerspiegelt. Exemplarisch sei hier das Item "Schwierigkeit scharf zu sehen" erwähnt, dass einen Haupteffekt Display [$F(1,36) = 28,662$; $p = .000$], einen Haupteffekt Zeitpunkt [$F(1,33) = 8,510$; $p = .000$] und Wechselwirkungen Display*Zeitpunkt [$F(1,33) = 7,668$; $p = .000$] aufweist. Hier zeigt sich aber auch eine Verringerung des Itemscores als Gewöhnung an die Displaytechnologie bei einem Vergleich zwischen 1. und 2. Messung mit dem HMD [$F(1,37) = 9,319$; $p = .004$]. Ebenso

erwähnenswert sind die Items, die nicht direkt der visuellen Ermüdung zuzuordnen sind: "Kopfschmerzen" zeigt höhere Werte beim HMD [$F(1,37) = 21.435$; $p = .000$], einen stärkeren Anstieg über die Zeit [$F(4,34) = 5.546$ $p = .002$] genauso wie "Nackenschmerzen" (Displayeffekt: [$F(1,37) = 22.000$; $p = .000$]; Wechselwirkung Display*Zeitpunkt [$F(4,34) = 5.341$; $p = .002$]). Und auch die geistige Ermüdung zeigt deutliche Effekte (Display [$F(1,37) = 7.956$; $p = .008$]; Zeitpunkt [$F(4,34) = 16.665$; $p = .000$]; Display*Zeitpunkt [$F(4,34) = 3.688$; $p = .013$]), wobei sich hier auch ein Gewöhnungseffekt im Vergleich zwischen 1. und 2. Messung zeigt [$F(1,37) = 4.175$; $p = .048$].

In einem abschließenden Interview bemängelten viele Probanden das Gewicht und den Tragekomfort des Kopfträgers. Die monokulare Darstellung vor nur einem Auge war für die meisten Probanden nach kurzer Eingewöhnungsphase kein Problem. Von 41 Probanden zog nur einer das HMD zum Arbeiten vor, während die anderen das Arbeiten mit dem Tablet-PC angenehmer fanden.

Diskussion

Die Ergebnisse zeigen allgemein schlechtere Leistung und höhere Beanspruchung beim Arbeiten mit einem HMD im Vergleich zum Tablet-PC. Auch in einer 2. Session und nach mehreren Stunden traten kaum Gewöhnungseffekte auf. Einige der berichteten Ergebnisse sind wahrscheinlich auch auf die Hardwareausführung des verwendeten HMDs zurückzuführen: Während die meisten aktuellen HMDs noch einen Kopfträger – ähnlich einem Fahrradhelm – verwenden, an dem die Technik befestigt ist, bringen die neuesten, gerade erst auf den Markt kommenden HMDs die gesamte Technik in einem etwas größeren Brillenmodell unter. Kopf- und Nackenschmerzen könnten also z. B. auch auf das Gewicht des Kopfträgers zurückzuführen sein. Auch muss erwähnt werden, dass im Rahmen der Laborstudie die Vorteile eines HMDs – die Mobilität und die frei bleibenden Hände – nicht zum Tragen kamen, was sich auf die hier vor allem berichteten subjektiven Kennwerte negativ auswirken könnte. Ebenso kann eine Gewöhnung an eine solche Technologie auch länger als 2 Sessions von je 4 Stunden dauern.

Die Vorgehensweise der lange andauernden Untersuchung und prozessbegleitenden Beanspruchungserhebung hat sich bewährt und konnte neben der höheren Beanspruchung beim HMD auch oft eine stärkere Zunahme über die Zeit nachweisen. Dies ist vor allem wichtig, wenn man das HMD als Arbeitsgerät verwenden will, um auf Basis dieser Erkenntnisse Empfehlungen für Pausenzeiten auszugeben. Für eine detailliertere Darstellung der Ergebnisse und weiterführende Diskussion sei deswegen hier auch auf den 2014 erscheinenden Forschungsbericht der BAuA verwiesen¹.

¹ <http://www.baua.de/de/Forschung/Forschungsprojekte/f2288.html>

Literatur

- Bangor, A. W. (2000): *Display technology and ambient illumination influences on visual fatigue at VDT Workstations*. Dissertation an der Virginia Polytechnic Institute and State University, USA.
- Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.
- Karrer, K., Glaser, C., Clemens, C., & Bruder, C. (2009). Technikaffinität erfassen – Der Fragebogen TA-EG. In: A. Lichtenstein, C. Stöbel & C. Clemens (Eds.), *Der Mensch als Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme* (pp. 196-201). Düsseldorf, Germany: VDI Verlag GmbH. Online verfügbar unter: <http://www.tu-berlin.de/fileadmin/f25/dokumente/8BWMMS/13.3-Karrer.pdf> [August 2013]
- Wille, M., Grauel, B. & Adolph, L. (2011) Methodische Vorgehensweise zur Beanspruchungsermittlung bei Head-Mounted Displays. In: *Reflexionen und Visionen der Mensch-Maschine-Interaktion – Aus der Vergangenheit lernen, Zukunft gestalten. 9. Berliner Werkstatt Mensch-Maschine-Systeme*. (S. 84-85). Düsseldorf: VDI Verlag, Fortschrittsberichte, Band 33. ISBN 978-3-18-303322-5.
- Zijlstra, F.R.H. (1993). *Efficiency in work behaviour: an approach for modern tools*. PHD Thesis, Delft University of Technology. Online verfügbar unter: <http://repository.tudelft.nl/view/ir/uuid:d97a028b-c3dc-4930-b2ab-a7877993a17f/> [August 2013].

Verringert Assistenz kognitive Fähigkeiten? - Auswirkungen unterschiedlicher Navigationsassistenz auf die räumliche Orientierung älterer Erwachsener in einer virtuellen Umgebung

Jörg Paschke¹ und Hartmut Wandke

Schlüsselwörter: räumliches Wissen, Navigationshilfen, ältere Erwachsene, Lernverläufe, virtuelle Umgebungen

Zusammenfassung

Inwiefern wirkt sich ein unterschiedliches Ausmaß an Assistenz auf den Erwerb räumlichen Wissens von Älteren aus? Verteilt in drei Experimentalgruppen, haben 60 ältere Erwachsene (61 - 85 Jahre) den Weg durch ein virtuelles Labyrinth gelernt. Dabei hatten sie entweder eine durchgehende Linie („volle“ Assistenz), einzelne Pfeile, die nur nach richtigen Abbiegungen sichtbar waren (Teilassistenz) oder keine zusätzliche Navigationshilfe zu Verfügung. Ohne Assistenz und mit Teilassistenz wurden signifikant mehr Landmarken wiedererinnert als mit voller Assistenz (Landmarkenwissen). Bezogen auf Zeit, Fehleranzahl und zurückgelegte Strecke im leeren Testlabyrinth ohne Landmarken und ohne Assistenz (Routenwissen) sowie beim Aufzeichnen der Landkarte des Labyrinths (Überblickswissen) fanden sich keine signifikanten Unterschiede zwischen den Bedingungen. Das Ausmaß an Assistenz hatte somit auf den Erwerb von Routen- sowie Überblickswissen bei älteren Erwachsenen keinen Einfluss. Landmarkenwissen geht hingegen durch das Nutzen eines hohen Assistenzmaßes zurück.

Einleitung

Moderne Technik ermöglicht Älteren mehr Unabhängigkeit im Alltag (Claßen, 2012; Burgard, Kiss & Wittmann, 2006; Rogers, Stronge & Fisk, 2005). Allerdings birgt z. B. der Gebrauch von Assistenzsystemen längerfristig das Risiko eines Abbaus, der durch diese ersetzt und folglich nicht genutzten Fähigkeiten oder Fertigkeiten. So wäre es denkbar, dass aufgrund der Nutzung von Navigationsgeräten räumliche Fähigkeiten verloren gehen (vgl. Lindenberger, Lövdén, Schellenbach, Li & Krüger 2008). Wenn dies zutrifft, hätten wir es hier mit einem Phänomen zu tun, das Bainbridge (1983) als eine von drei Ironien der Automatisierung, allerdings im Kontext großtechnischer Anlagen, bezeichnet.

Es ist von großer Bedeutung, (gute) räumliche Fähigkeiten zu erhalten. Räumliche Fähigkeiten nehmen altersbedingt ab, was zu Restriktionen im alltäglichen Leben führen kann (Young & Bunce, 2011; Rogers et al., 2005; Simoes, 2003; Lin-

¹ Humboldt-Universität zu Berlin, Institut für Psychologie, Ingenieurpsychologie/Kognitive Ergonomie, Unter den Linden 6, 10099 Berlin; joerg.paschke@hu-berlin.de, hartmut.wandke@psychologie.hu-berlin.de

denberger, 2000). So kann es sein, dass ältere Erwachsene aufgrund verminderter Leistungen bei der mentalen Rotation, der Bildung kognitiver Karten oder beim Wegfinden u. a. Schwierigkeiten bei der Orientierung im Raum haben und deswegen z. B. nur noch eingeschränkt Auto fahren (Simoes, 2003). Zudem sind räumliche Fähigkeiten eine Voraussetzung für die effiziente Verarbeitung bildhaft präsentierter Information (vgl. Souvignier, 2001) und für den Umgang mit den neuen Technologien wie z. B. Computern (Rogers et al., 2005).

Beim Erwerb räumlichen Wissens in einer neuen Umgebung werden drei Komponenten unterschieden: Landmarken-, Routen- und Überblickswissen (Thorndyke, 1980; Thorndyke & Hayes-Roth, 1982). Landmarken sind saliente Objekte in einer Umgebung, die zum Navigieren verwendet werden können. Routenwissen ist zum einen das Wissen über Anordnung von Landmarken, zum anderen eine Sequenz direktonaler Entscheidungen. Überblickswissen ist eine kartenähnliche, zweidimensionale Repräsentation der Umgebung (vgl. Jansen, Schmelter & Heil, 2010; vgl. Hamburger, Röser, Bukow & Knauff, 2012). Landmarken unterstützen Erwerb und Abruf von Routenwissen, sind dafür aber nicht notwendig (Jansen-Osman, 2002; vgl. Mallot, 2006).

Der Erwerb räumlichen Wissens fällt mit zunehmendem Alter schwerer. Jansen et al. (2010) untersuchten Erwachsene jüngeren (20 – 30 Jahre), mittleren (40 – 50 Jahre) und höheren (60-70) Alters in einer virtuellen Umgebung (VU). Ältere brauchen länger, eine neue Route zu lernen und zeigen geringere Leistungen beim Wiedererinnern von Landmarken und Zeichnen einer Karte der Umgebung.

Es gibt auch Alterseffekte beim Unterstützungspotential unterschiedlicher Assistenz. Bei Schellenbach, Lövdén, Verrel, Krüger und Lindenberger (2010) profitierten in einer virtuellen Navigationsaufgabe die älteren Versuchspersonen (68 - 77 Jahre) gleichermaßen wie die jüngeren (21 - 28 Jahre) von einem durchgehenden Linie als Navigationshilfe. Beide profitierten auch von einer ständig angezeigten Überblickskarte, die Älteren allerdings viel weniger als die Jüngeren.

Ziel der vorliegenden Studie war es, zu überprüfen, ob ein Gebrauch von Navigationsassistenz tatsächlich negative Auswirkungen auf Aspekte räumlicher Fähigkeiten bei Älteren hat. Dafür nutzen wir die VU von Jansen et al. (2010) und entwarfen zwei weitere Versionen mit Navigationshilfen, die unterschiedlich hoch das Wegfinden unterstützten. Dabei stellte die sogenannte Teilassistenz eine stark reduzierte Variante der „vollen“ Assistenz dar. Untersucht wurde, ob ein unterschiedliches Ausmaß an Navigationsassistenz, d. h. das Vorliegen unterschiedlich verfügbarer Hilfsinformationen zum Wegfinden, auch mit korrespondierendem, unterschiedlichem räumlichen Wissen, der neu gelernten(virtuellen) Umgebung einhergeht. Wir erwarteten mit steigendem Assistenzausmaß, ein geringeres Landmarken-, Routen- sowie Überblickswissen.

Methode

Insgesamt 60 Versuchspersonen im mittleren Alter von 69.57 Jahren (Spanne: 61 - 85 Jahre; SD=6.07) nahmen an der Studie teil. Die anfallende Stichprobe bestand zur Hälfte aus Frauen. Die Teilnehmenden erhielten eine Aufwandsentschädigung von 10 EUR. Eine Experimentalsitzung dauerte ca. eine Stunde.

In dem Experiment wurde eine VU genutzt, die mit dem Software-Programm Gamestudio A8 (Conitec Datensysteme GmbH, Dieburg, Deutschland) erstellt wurde. Das Programm wurde auch als Experimentalsoftware zur Präsentation und Datenaufzeichnung verwendet. Das Experiment lief auf einem Laptop-Computer (Intel-Core i5, 2,3 GHz) mit einer NVIDIA GeForce 410M Grafikkarte. An diesem war ein 23 Zoll Widescreen-Flachbildschirm als Experimentalbildschirm angeschlossen. Die Versuchspersonen saßen so, dass ihre Augen ca. 60 cm von diesem entfernt waren. Die Bildschirmauflösung betrug 1366 mal 768 Pixel.

Der Aufbau der VU basierte auf dem Labyrinth der Studie von Cohen und Schuepfer (1980), welches von Jansen und Mitarbeitern als VU nachgestaltet wurde (z. B. Jansen-Osmann, 2002; Jansen et al., 2010). In dem virtuellen Labyrinth aus mehreren Gängen und Sackgassen gab es nur einen Weg zum Ziel. Darin verteilt standen verschiedene Landmarken mit unterschiedlicher Funktion: Je sechs Landmarken standen neben richtigen bzw. falschen Abbiegungen oder an keiner Abbiegung (siehe Abb. 1). Da Cohen und Schuepfer sowie Jansen et al. Kinder untersuchten, wurden Stofftiere als Landmarken verwendet. Die Navigation in der VU erfolgte in Egoperspektive.

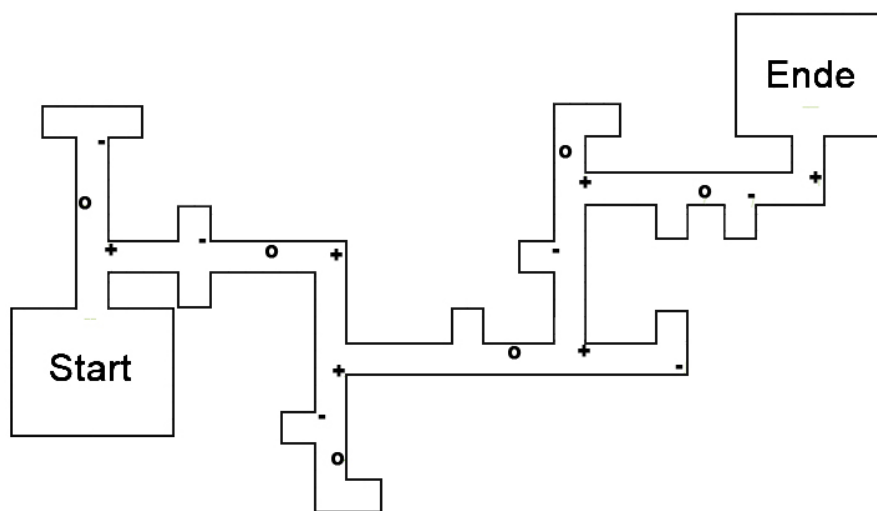


Abb. 1: Aufsicht auf das virtuelle Labyrinth; „+“ kennzeichnet Landmarken neben richtigen Abbiegungen, „-“ neben falschen Abbiegungen, „o“ neben keiner Abbiegung

In einem Between-subjects-Design gab es drei unabhängige Stichproben mit jeweils unterschiedlichem Ausmaß an Navigationshilfe (siehe Abb. 2). Bei der vollen Assistenz gab es eine vom Anfang bis zum Ende durchgehende gelbe Linie auf dem Boden des Labyrinths (siehe Abb. 2 C), ähnlich Schellenbach et al. (2010). Die Linie war zusätzlich mit dicht aufeinanderfolgenden Pfeilen versehen, welche

die Richtung zum Ziel indizierten. Bei der Teilassistenz zeigten einzelne Pfeile nach den richtigen Abbiegungen den Weg an, die erst bei Blick in den jeweiligen Korridor sichtbar waren (siehe Abb. 2 B). Eine weitere Experimentalgruppe erhielt keine derartige Navigationshilfe (siehe Abb. 2 A). Die Versuchspersonen erfuhren vor dem Experiment nichts über die Landmarken. Den Gruppen mit Assistenz wurde die Funktion der jeweiligen Navigationshilfe kurz erläutert.

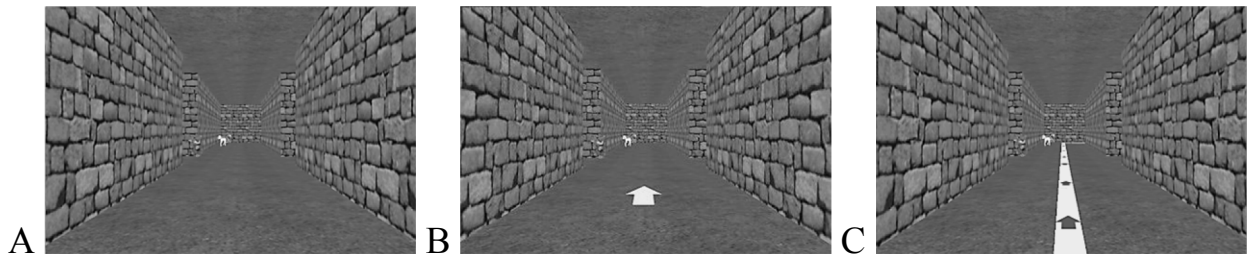


Abb. 2: Blick in das virtuelle Labyrinth A) ohne Assistenz, B) mit Pfeilen nach richtigen Abbiegungen als Teilassistenz und C) mit durchgehender Linie als Vollassistenz.

Vor dem Experiment absolvierten die Versuchspersonen ein Tutorial zur Navigation mit den vier Richtungstasten (nach vorne bzw. hinten bewegen, nach links bzw. rechts drehen) der Response Box und navigierten anschließend durch ein kleines Probelabyrinth (eine Linksabbiegung, keine Sackgassen). Danach konnten sie auf Wunsch das Navigieren frei weiterüben. Anhand der Zeit, die benötigt wurde, das Probelabyrinth zu durchlaufen, wurden die Versuchspersonen einer der drei Experimentalgruppen zugeordnet.

In der Lernphase hatten die Versuchspersonen die Aufgabe, sechsmal nacheinander zum Ende des virtuellen Labyrinths zu finden. Bei Jansen et al. (2010) reichte den Älteren (60 – 70 Jahre) diese Durchgangsanzahl, um die richtige Route durch das Labyrinth zu lernen (benötigte Lerndurchgänge bei Jansen et al., um danach zweimal fehlerfrei durch den Gang zu laufen: $M=6.00$, $SE=0.56$).

Das Vorgehen in der Testphase orientierte sich an dem Paradigma von Jansen et al. (2010) bzw. Cohen und Schuepfer (1980). Der Ablauf der Testphase wurde vor dem Experiment nicht angekündigt. Direkt nach dem sechsten Lerndurchgang mussten die Versuchspersonen einmal durch das leere Labyrinth - ohne Landmarken und ohne Navigationsassistenz - bis zum Ende laufen. Durch automatische Aufzeichnung eines Logfiles wurden Zeiten, zurückgelegte Strecke sowie Fehler (Blicken in eine Sackgasse oder Navigieren in einen vom richtigen Weg abweichenden Teil eines Ganges) während der einzelnen Durchgänge in der Lernphase sowie im Testdurchlauf (Routenwissen) gemessen. Das Landmarkenwissen wurde auch im leeren Labyrinth getestet, indem die Versuchspersonen an vermuteten Lokalisationen von Landmarken anhielten und diese verbal benannten. Der Versuchsleiter saß dabei neben der Versuchsperson und markierte verdeckt richtige Lösungen auf einer Karte des Labyrinths mit Landmarken. Zudem wurde gefragt, welche Landmarkenidentitäten noch wiedererinnert wurden. Um Überblickswissen zu erfassen, sollten die Versuchspersonen eine Karte des Labyrinths aufzeichnen. An-

hand der Karten wurden mittels zweier Rater ein sogenannter Map Correctness Score gebildet, bei dem es maximal 12 Punkte zu vergeben gab und der sich auf wichtige Merkmale des Labyrinths, wie z. B. die Anzahl der Sackgassen, bezieht (zum Vorgehen der Punktevergabe: vgl. Jansen et al., 2010, S. 56). Der so ermittelte Map Correctness Score wies eine gute interne Konsistenz auf (Cronbachs $\alpha=.86$).

Mit einem soziodemographischen Fragebogen wurden Alter, Geschlecht und höchster Bildungsabschluss der Versuchspersonen erfasst. Zudem beantworteten sie Fragen über Dauer (in Jahren) und Intensität (in Stunden pro Woche) zur Nutzung von Computern sowie von Computerspielen. Zum Erfassen eines Aspekts räumlich-kognitiver Fähigkeiten - der mentalen Rotation - wurde der Untertests 7 „Mentales räumliches Rotieren“ aus dem Leistungsprüfsystem für 50-90jährige (LPS 50+; Sturm, Willmes & Horn, 1993) erhoben. Die visuelle Kurzzeitgedächtnisleistung wurde über den Bildertest aus dem Nürnberger-Alters-Inventar (NAI; Oswald & Fleischmann, 1986) ermittelt.

Ergebnisse

Vor der statistischen Auswertung wurden in den abhängigen Variablen (AVn) Ausreißerwerte, die über drei Standardabweichungen lagen ermittelt. Anstatt diese zu exkludieren, wurden sie ersetzt durch den Kriteriumswert (Mittelwert plus bzw. minus 3 Standardabweichungen), wie u. a. von Field (2009) vorgeschlagen, da sehr hohe bzw. niedrige Werte als valider Teil der Stichprobe gesehen wurden und nicht als Verzerrung durch zufallsbedingt externe Faktoren. Auf diese Weise wurde in den drei Teilstichproben maximal ein Ausreißer je AV ersetzt. Um die Effekte des Assistenzmaßes auf das räumliche Wissen zu untersuchen, wurden zunächst univariate Varianzanalysen (ANOVAs) berechnet mit der Experimentalbedingung als Zwischensubjektfaktor. Die α -Wahrscheinlichkeit wurde auf 5 % festgelegt.

Vergleichbarkeit der Experimentalgruppen: Die drei Experimentalgruppen unterschieden sich weder in den erhobenen soziodemographischen Daten, in der Erfahrung mit Computern bzw. Computerspielen, noch in den Leistungen bei den kognitiven Tests.

Landmarkenwissen: Bei der Lokalisation von Landmarken fand sich ein signifikanter Effekt der Bedingung, $F(2,57)=7.34$, $p<.01$, $\eta^2=.20$ (siehe Abb. 3). Geplante Kontraste zeigten auf, dass wenn in der Lernphase volle Assistenz ($M=0.61$, $SE=0.31$) zur Verfügung stand, weniger Landmarken wiedererinnert wurden im Vergleich zu keiner Assistenz ($M=2.65$, $SE=0.48$), $t(32.53)=3.55$, $p<.001$ (1-seitig), $d=1.12$. Dies gilt ebenso für die volle Assistenz im Vergleich zur Teilassistenz ($M=1.80$, $SE=0.31$), $t(38.00)=2.69$, $p<.01$ (1-seitig), $d=0.85$. Zwischen den Bedingungen ohne Assistenz und Teilassistenz, $t(32.57)=1.48$, $p=.07$ (1-seitig), $d=0.47$, gab es zwar keinen signifikanten Unterschied beim Landmarkenwissen, die Effektgröße deutet aber darauf hin, dass dies möglicherweise an der zu kleinen Stichprobengröße liegt.

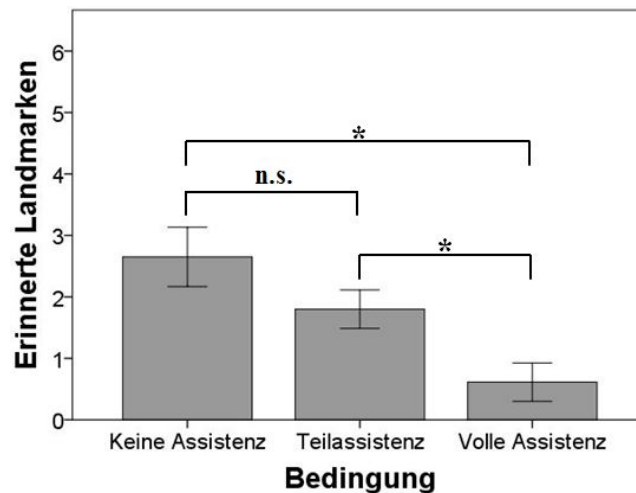


Abb. 3: Mittelwerte richtiger Lokalisationen von Landmarken in Abhängigkeit der Bedingung; „*“ kennzeichnet einen signifikanten, „n.s.“ einen nicht signifikanten Unterschied zwischen den Bedingungen; Fehlerbalken kennzeichnen Standardfehler.

Routenwissen: Bezogen auf Zeit, $F(2,57)=0.70$, $p=.50$, $\eta^2=.02$, zurückgelegte Strecke, $F(2,57)=0.26$, $p=.77$, $\eta^2=.01$, sowie Fehleranzahl, $F(2,57)=0.06$, $p=.95$, $\eta^2=.00$ im leeren Labyrinth des Testdurchgangs fanden sich keine signifikanten Unterschiede zwischen den Bedingungen.

Überblickswissen. Es gab keinen signifikanten Effekt der Bedingung beim Map Correctness Score, $F(2,57)=0.72$, $p=.49$, $\eta^2=.02$. Das Überblickswissen war bei allen Assistenzbedingungen gleich hoch.

Lernverläufe: Um die einzelnen Lernverläufe des Routenwissens in der VU zu untersuchen, wurden Varianzanalysen mit Messwiederholung durchgeführt mit der Assistenzbedingung als Zwischensubjektfaktor sowie mit der Zeit, das virtuelle Labyrinth zu durchqueren, der zurückgelegten Strecke sowie der Fehleranzahl in den sechs einzelnen Durchgängen als AVn. Es gab einen signifikanten Haupteffekt der Lerndurchgänge auf die Zeit, $F(1.68, 95.60)=27.93$, $p<.001$. Die Interaktion von Bedingung und Lerndurchgang war nicht signifikant, was zeigt, dass der Verlauf des Lernens über die wiederholten Durchgänge in allen drei Assistenzbedingungen ähnlich war. Es gab auch einen Haupteffekt der Bedingung, $F(2, 57)=8.02$, $p<.01$. Bonferroni Post-hoc-Tests des Haupteffekt der Bedingung zeigten, dass mit voller Assistenz weniger Zeit benötigt wurde, den Weg durch das Labyrinth zu finden, als ohne Assistenz ($p<.01$) und mit Teilassistenz ($p<.05$). Die Bedingungen ohne Assistenz und Teilassistenz waren hingegen gleich schnell ($p=.85$; siehe Abb. 4).

Es gab einen signifikanten Haupteffekt der Lerndurchgänge auf die zurückgelegte Strecke, $F(1.28, 72.88)=4.79$, $p<.05$. Es gab keine signifikante Interaktion. Ferner gab es auch einen Haupteffekt der Bedingung, $F(2, 57)=4.95$, $p<.05$. Bonferroni Post-hoc-Tests des Haupteffekt der Bedingung zeigten, dass ohne Assistenz eine längere Strecke im Labyrinth zurückgelegt wurde als mit voller Assis-

tenz ($p < .01$). Mit Teilassistentz gab es keine Unterschiede in der zurückgelegten Strecke zur vollen Assistentz ($p = .34$) sowie ohne Assistentz ($p = .39$; siehe Abb. 4).

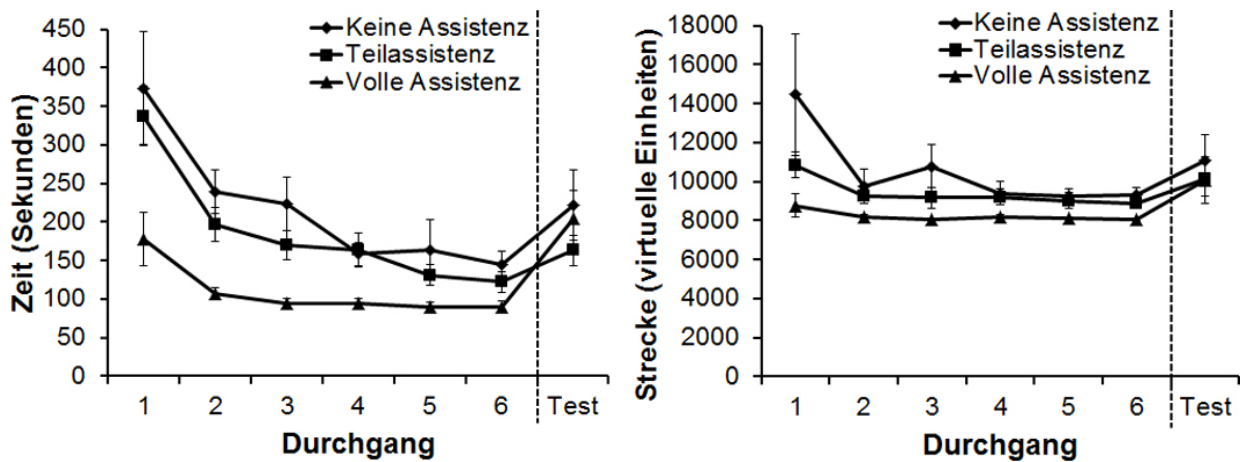


Abb. 4: Benötigte Zeit (links) und zurückgelegte Strecke (rechts) je Lern- (1 - 6) sowie im Testdurchgang beim Durchqueren des virtuellen bzw. leeren Labyrinths in den unterschiedlichen Assistenzbedingungen. Fehlerbalken kennzeichnen Standardfehler.

Es gab einen Haupteffekt der Lerndurchgänge auf die Fehleranzahl, $F(1.87, 106.38) = 24.22$, $p < .001$. Es gab eine signifikante Interaktion zwischen Lerndurchgängen und Bedingung, $F(3.73, 106.38) = 4.82$, $p < 0.01$. Es gab einen Haupteffekt der Bedingung, $F(2, 57) = 44.55$, $p < .001$. Bonferroni Post-hoc-Tests des Haupteffekt der Bedingung zeigten, dass mit voller Assistentz weniger Fehler gemacht wurden als ohne Assistentz ($p < .001$) und mit Teilassistentz ($p < .001$). Ohne und mit Teilassistentz wurden gleich viele Fehler gemacht ($p = 1.00$; siehe Abb. 5).

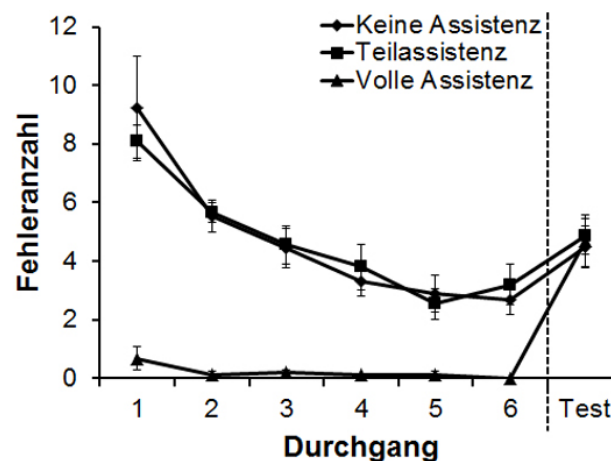


Abb. 5: Fehleranzahl je Lern- (1 - 6) sowie im Testdurchgang in den unterschiedlichen Assistenzbedingungen. Fehlerbalken kennzeichnen Standardfehler.

Diskussion

Unsere Ergebnisse widerlegten einen Teil unserer Annahmen, dass mit steigendem Ausmaß an Navigationshilfe, grundsätzlich ein verringerter Erwerb an räumlichem Wissen bei älteren Erwachsenen einhergeht. Das Ausmaß an Assistenz hat auf den

Erwerb von Routen- sowie Überblickswissen keinen Einfluss. Allerdings zeigte sich ein geringeres Landmarkenwissen bei der vollen Assistenz, im Vergleich zur fehlenden oder Teilassistenz. Hingegen unterschied sich das Landmarkenwissen nicht zwischen den Bedingungen ohne Assistenz und mit Teilassistenz. Während explizit verfügbares Landmarkenwissen durch das Nutzen voller Assistenz zurückgeht, gibt es keine Einschränkungen bei implizitem Wissen, das der Routenwahl zugrunde liegt.

Die höhere kognitive Entlastung durch eine ständig präsente Navigationshilfe, könnte zu einer Kompensation beim Erwerb von Routen- und Überblickswissens geführt haben. Nach Schellenbach et al. (2010) ist die von uns verwendete volle Assistenz eine Navigationshilfe mit niedriger kognitiver Beanspruchung. Die Teilassistenz hätte demnach eine höhere kognitive Beanspruchung. Anders als bei voller Assistenz, standen dabei die richtungsanzeigenden Informationen nicht permanent zur Verfügung, sondern mussten beim Explorieren des Weges in der VU (wieder-) entdeckt werden. Folgt man der Interpretation von Schellenbach et al., könnte auch bei uns, aufgrund der durchgehenden Linie, die kognitive Beanspruchung von Navigation, im Vergleich zu der Bedingung ohne Assistenz, stark reduziert worden sein. Dies träfe bei der Teilassistenz nur im geringen Maße zu. Während sich für das Wegfinden ohne Assistenz und mit Teilassistenz eingehender mit der Umgebung auseinandergesetzt werden musste, könnten die freien kognitiven Ressourcen aufgrund der vollen Assistenz dazu genutzt worden sein, sich Routen- und Überblickswissen im gleichem Masse anzueignen.

Das Paradigma von Jansen et al. (2010) bzw. Cohen und Schuepfer (1980) eignet sich, die drei Komponenten räumlichen Wissens zu erfassen. So wird z. B. in gesonderten Testdurchgängen im leeren Labyrinth Routenwissen getrennt vom Landmarkenwissen erfasst. Es könnte allerdings sein, dass die drei Komponenten nicht unabhängig voneinander sind. Es ist z. B. nicht auszuschließen, dass Landmarken auch implizit gelernt werden, die dann zwar nicht explizit an der richtigen Position abrufbar sind, dennoch den Erwerb und den Abruf von Routenwissen unterstützen. Ein Hinweis darauf ist, dass auf Nachfrage meistens mehr Landmarkenidentitäten frei erinnert, als im Test richtig lokalisiert wurden. Diese Problematik könnte in nachfolgenden Experimenten geklärt werden.

In den Lernverläufen über die sechs Lerndurchgänge zeigte sich eine Überlegenheit der vollen Assistenz bezüglich ihrer Unterstützungsfunktion. Mit voller Assistenz wurde in der Lernphase das Labyrinth schneller durchlaufen und wurden sehr wenige Fehler gemacht im Vergleich zu den Bedingungen ohne Assistenz und mit Teilassistenz. Allerdings schnitt die Teilassistenz bei der zurückgelegten Strecke in den Lerndurchgängen nicht schlechter ab als die volle Assistenz. Folglich bringt die volle Assistenz die älteren Erwachsenen schneller und fehlerfreier zum Ziel als die Teilassistenz. Beide Navigationshilfen halten diese aber im Vergleich zur fehlenden Assistenz eher auf dem richtigen Weg. Dabei kann man insgesamt einen typischen Lernverlauf in jeder Bedingung erkennen (Heathcote, Brown & Mewhort, 2000), der sich auch bei älteren Erwachsenen zeigt. Eine vergleichbar

fallende Lernkurve für die benötigte Zeit und zurückgelegte Strecke in einem virtuellen Labyrinth (ohne Navigationshilfe) ermittelten auch Moffat, Zonderman und Resnick (2001) für ältere Erwachsene (66 – 91 Jahre).

Limitationen: In der vorliegenden Studie wurden nur kurzfristige Effekte von Assistenzformen untersucht. Der als Ironie der Automatisierung (Bainbridge, 1983) bekannte Kompetenzverlust tritt langfristig auf. Der signifikante Unterschied beim Erinnern der Landmarken kann als kurzfristiger Effekt im Sinne von „working storage“ (Bainbridge, 1983) oder „situation awareness“ (Endsley, 1995) interpretiert werden. Wir wissen nicht, ob räumliche Orientierung durch volle Assistenz langfristig doch vermindert wird.

Die nicht eingetretenen Effekte könnten auch dadurch bedingt sein, dass ältere Versuchspersonen besonders gewissenhaft und zuverlässig agieren. Durch eine Wiederholung des Experiments mit jüngeren Versuchspersonen könnte diese Annahme geprüft werden.

Ob ältere Menschen sich generell weniger auf Assistenzfunktionen verlassen und dazu neigen, durch eigene Handlungen Aufgaben zu bearbeiten, müsste durch einen Wechsel der Aufgabe und der Assistenzfunktion geprüft werden.

Sollte sich der von uns gefundene Effekt auch in anderen Situationen (z. B. in komplexeren Umgebungen), bei anderen Assistenzformen und über einen längeren Zeitraum zeigen, so wären die eingangs erwähnten Befürchtungen zur Nutzung von Assistenz ungerechtfertigt.

Danksagung

Unser besonderer Dank geht an Frau Prof. Dr. Petra Jansen, Leiterin des Instituts für Sportwissenschaft an der Universität Regensburg, die uns freundlicherweise den Grundaufbau der virtuellen Umgebung zur Verfügung gestellt hat.

Literatur

- Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19 (6), 775–779.
- Burgard, E., Kiss, M. & Wittmann, M. (2006). Gerontotechnik. In W. D. Oswald, U. Lehr, C. Sieber & J. Kornhuber (Hrsg.), *Gerontologie. Medizinische, psychologische und sozialwissenschaftliche Grundbegriffe* (3. Aufl., S. 199–204). Stuttgart: Kohlhammer.
- Claßen, K. (2012). Technik im Alltag. In H.-W. Wahl, C. Tesch-Römer & J. P. Ziegelmann (Hrsg.), *Angewandte Gerontologie. Interventionen für ein gutes Altern in 100 Schlüsselbegriffen* (2. Aufl., S. 499–506). Stuttgart: Kohlhammer.
- Cohen, R. & Schuepfer, T. (1980). The representation of landmarks and routes. *Child Development*, 51, 1065–1071.

- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors* 37 (1), 32–64.
- Field, A. (2009). *Discovering statistics using SPSS* (3. Aufl.). London: SAGE.
- Hamburger, K., Röser, F., Bukow, G. C. & Knauff, M. (2012). Der Weg als Ziel. Virtuelle Umgebungen und räumlicher Wissenserwerb. In G. C. Bukow, J. Fromme & B. Jörissen (Hrsg.), *Raum, Zeit, Medienbildung* (S. 173–193). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Heathcote, A., Brown, S. & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, 7 (2), 185–207.
- Jansen, P., Schmelter, A. & Heil, M. (2010). Spatial knowledge acquisition in younger and elderly adults. *Experimental Psychology*, 57 (1), 54–60.
- Jansen-Osmann, P. (2002). Using desktop virtual environments to investigate the role of landmarks. *Computers in Human Behavior*, 18, 427–436.
- Lindenberger, U. (2000). Intellektuelle Entwicklung über die Lebensspanne: Überblick und ausgewählte Forschungsbrennpunkte. *Psychologische Rundschau*, 51 (3), 135–145.
- Lindenberger, U., Lövdén, M., Schellenbach, M., Li, S.-C. & Krüger, A. (2008). Psychological principles of successful aging technologies: A mini-review. *Gerontology*, 54, 59–68.
- Mallot, H. A. (2006). Raumorientierung und kognitive Karten. In H.-O. Karnath & P. Thier (Hrsg.), *Neuropsychologie* (S. 153–159). Berlin: Springer.
- Moffat, S. D., Zonderman, A. B. & Resnick, S. M. (2001). Age differences in spatial memory in a virtual environment navigation task. *Neurobiology of Aging*, 22, 787–796.
- Oswald, W. D. & Fleischmann, U. M. (1986). *Nürnberger-Alters-Inventar (NAI). Testkasten & NAI-Kurzmanual*. Nürnberg: Universität Erlangen-Nürnberg.
- Rogers, W. A., Stronge, A. J. & Fisk, A. D. (2005). Technology and Aging. *Reviews of Human Factors and Ergonomics*, 1 (1), 130–171.
- Thorndyke, P. W., & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology*, 14, 560–589.
- Thorndyke, P. W. (1980). *Performance models for spatial and locational cognition* (Technical Report R-2676-ONR). Santa Monica, CA: Rand.
- Schellenbach, M., Lövdén, M., Verrel, J., Krüger, A. & Lindenberger, U. (2010). Sensorimotor-cognitive couplings in the context of assistive spatial navigation for older adults. *GeroPsych*, 23 (2), 69–77.
- Simoes, A. (2003). The cognitive training needs of older drivers. *Recherche Transports Sécurité*, 79, 145–155.
- Souvignier, E. (2001). Training räumlicher Fähigkeiten. In K. J. Klauer (Hrsg.), *Handbuch kognitives Training* (2. Aufl., S. 293–319). Göttingen: Hogrefe.

- Sturm, W., Willmes, K. & Horn, W. (1993). *Leistungsprüfsystem für 50-90jährige (LPS 50+)*. Göttingen: Hogrefe.
- Young, M. S. & Bunce, D. (2011). Driving into the sunset: Supporting cognitive functioning in older drivers. *Journal of Aging Research*, 2011, 1–6.

Autorenverzeichnis

Adolph, Lars	Bundesanstalt für Arbeitsschutz und Arbeitsmedizin
Adolphs, Nicolai	Klinik für MKG-Chirurgie Zentrum für rekonstruktive und plastisch-ästhetische Gesichtschirurgie, Berlin, Deutschland
Altendorf, Eugen	RWTH Aachen, Lehrstuhl und Institut für Arbeitswissenschaft
Antons, Jan-N.	Technische Universität Berlin, Quality und Usability Lab
Arndt, Sebastian	Technische Universität Berlin
Auricht, Maik	Technische Universität Berlin
Backhaus, Nils	Technische Universität Berlin
Balaut, Magali	Technische Universität Berlin, GRK Prometei
Baltzer, Marcel	RWTH Aachen, Lehrstuhl und Institut für Arbeitswissenschaft
Baumann, Martin	Deutsches Zentrum für Luft- und Raumfahrt
Behrenbruch, Kay	Universität Kassel
Bengler, Klaus	Technische Universität München, Lehrstuhl für Ergonomie
Berdjag, Denis	LAMIH, The University of Valenciennes and Hainaut-Cambresis, France
von Bernstorff, Charlotte	Technische Universität Berlin, Zentrum für Mensch Maschine Systeme (ZMMS)
Bleyer, Anne	Technische Universität Berlin, Fachgebiet Arbeits-, Ingenieur- und Organisationspsychologie
Bock, Juliane	Technische Universität Berlin, Institut für Psychologie und Arbeitswissenschaft, Fachgebiet Mensch-Maschine-Systeme
Brandenburg, Stefan	Technische Universität Berlin
Brandstedter, Peter	Technische Universität Berlin
Bremer, Andreas	Institut für Arbeitsforschung an der TU Dortmund (IfADo)
Bretschneider, Hannes	Technische Universität Dresden, Professur für Verkehrspsychologie
Brüngger, Jonas	Fachhochschule Nordwestschweiz, Schweiz
Bullinger, Angelika	Technische Universität Chemnitz, Professur Arbeitswissenschaft

C.	und Innovationsmanagement
Bürglen, Juliane	Technische Universität Berlin, Quality und Usability Lab
Burmester, Michael	Hochschule der Medien (HdM), Stuttgart. Information Experience and Design Research Group
Buyer, Sven	Hochschule Heilbronn, Deutschland
Carbon, Claus C.	Otto-Friedrich-Universität Bamberg, Lehrstuhl für Allgemeine Psychologie
Debernard, Serge	University of Valenciennes, France
Deutsch, Daniela	Technische Universität Dresden, Professur für Verkehrspsychologie
Dittrich, Elisabeth	Technische Universität Berlin
Dohle, Christian	Abteilung für Neurologische Rehabilitation, MEDIAN Klinik Berlin-Kladow, Centrum für Schlaganfallforschung Berlin, Charité Universitätsmedizin Berlin, Professur für Rehabilitationswissenschaften, Universität Potsdam
Dannehl, Susanne	Technische Universität Berlin, Fachgebiet Medizintechnik
Domhardt, Michael	Universität Kassel
Doria, Laura	Technische Universität Berlin
Dziennus, Marc	Deutsches Zentrum für Luft- und Raumfahrt
Eckstein, Lutz	Institut für Kraftfahrzeuge, RWTH Aachen University
Eggert, Frank	Technische Universität Braunschweig, Institut für Psychologie, Abteilung Psychologische Methodenlehre und Biopsychologie
Engelbrecht, Klaus-P.	Quality and Usability Lab, Telekom Innovation Laboratories, Technische Universität Berlin
Feige, Tim	Technische Universität Dresden, Professur für Verkehrspsychologie
Ferreira dos Santos, Luara	Technische Universität Berlin
Fischer, Katrin	Fachhochschule Nordwestschweiz CH, Schweiz
Flemisch, Frank Ole	Ergonomie und Mensch-Maschine-Systeme, Fraunhofer Institut
Floris, Jérôme	LAMIH, CNRS UMR 8201, Université de Valenciennes et du Hai-

	naut Cambr�sis, France
Fortmann, Florian	Human Machine Interaction, OFFIS
Freude, Gabriele	Bundesanstalt f�r Arbeitsschutz und Arbeitsmedizin
Freytag, Sarah C.	Technische Universit�t Berlin, Institut f�r Psychologie und Arbeitswissenschaft, Fachgebiet Mensch-Maschine-Systeme
Fricke, Nicola	Deutsches Zentrum f�r Luft- und Raumfahrt
G�rtner, Kathrin	Fachhochschule Nordwestschweiz, Schweiz
Gellot, Fran�ois	CR�STIC URCA, UFR Sciences Exactes et Naturelles, Moulin de la Housse, France
Gong, Xin Guang	Technische Universit�t Berlin
Grattenthaler, Heidi	Interdisziplin�res Zentrum f�r Verkehrswissenschaften an der Universit�t W�rzburg (IZVW)
Grauel, Britta	Bundesanstalt f�r Arbeitsschutz und Arbeitsmedizin
Grauert, Josephine	Technische Universit�t Berlin, Kognitionspsychologie und Kognitive Ergonomie
Gross, Alice	Technische Universit�t Berlin
Grossenbacher, Peter	Schweizerische Bundesbahnen SBB, Schweiz
G�nzler, Torsten	Technische Universit�t Berlin
Haiduk, Michael	Technische Universit�t Braunschweig, Institut f�r Psychologie, Abteilung Psychologische Methodenlehre und Biopsychologie
Halbr�gge, Marc	Technische Universit�t Berlin, Deutschland
Hasselberg, Andreas	Deutsches Zentrum f�r Luft- und Raumfahrt e.V. (DLR)
Heesen, Matthias	Fraunhofer-Institut f�r Kommunikation, Informationsverarbeitung und Ergonomie FKIE
Helmbrecht, Magnus	Technische Universit�t M�nchen
Heyne, Fanny	Technische Universit�t Dresden
Hillebrand, Axel	EADS Innovation Works, Hamburg
Hoffmann, Holger	Universit�t Kassel, Fachgebiet Wirtschaftsinformatik

Hoffmeister, Bodo	Klinik für MKG-Chirurgie Zentrum für rekonstruktive und plastisch-ästhetische Gesichtschirurgie
Hoppe, Florian	Technische Universität Berlin, Institut für Psychologie und Arbeitswissenschaft, Fachgebiet Mensch-Maschine Systeme
Horn, Maximilian	Technische Universität Braunschweig, Institut für Psychologie, Abteilung Psychologische Methodenlehre und Biopsychologie
Israel, Johann H.	Fraunhofer-Institut für Produktionsanlagen- und Konstruktionstechnik (IPK)
Issing, Matthias	medneo GmbH, Berlin
Ivanova, Ekaterina	Technische Universität Berlin
Joeres, Fabian	Technische Universität Berlin
Johnsen, Annika	Technische Universität Berlin, Fachgebiet Mensch-Maschine-Systeme
Jonker, Catholijn	Delft University of Technology Faculty of Electrical Engineering, Mathematics and Computer, Netherlands
Josten, Johanna	Institut für Kraftfahrzeuge, RWTH Aachen University
Jumisko-Pyykkö, Satu	Tampere University of Technology, Finland
Junge, Marko	Volkswagen AG
Karrer-Gauß, Katja	Technische Universität Berlin
Keeve, Erwin	Fraunhofer Institute for Production Systems and Design Technology, Berlin
Kelsch, Johann	Deutsches Zentrum für Luft- und Raumfahrt, Institut für Verkehrssystemtechnik
Kieselmann, Olga	Angewandte Informationssicherheit, OFFIS
Kolbig, Maria	BMW Group, München
Kolski, Christophe	University of Valenciennes, France
Kopal, Nils	Universität Kassel, Fachgebiet Angewandte Informationssicherheit
Körber, Moritz	Lehrstuhl für Ergonomie, Technische Universität München
Kraft, Marc	Technische Universität Berlin, Fachgebiet Medizintechnik

Krott, Tatiana	Hochschule für Medien, Kommunikation und Wirtschaft (HMKW), Campus Köln
Krüger, Jörg	Forschungsgruppe Rehabilitationsrobotik (IPK/TU Berlin), Abteilung Automatisierungstechnik, Fraunhofer Institut für Produktionsanlagen und Konstruktionstechnik (IPK), Berlin
Kuz, Sinem	Institut für Arbeitswissenschaft der RWTH Aachen
Lange, Michael	The Tower Company GmbH, Langen
Lebrun, Yoann	PRL-CCI Grand Hainaut, France
Leimeister, Jan M.	Universität Kassel, Fachgebiet Wirtschaftsinformatik
Lekkala, Jukka	Department of Automation Science and Engineering, Tampere University of Technology, Finland
Lepreux, Sophie	PRL-CCI Grand Hainaut, France
Ley, Daniel	Fraunhofer FKIE
Liu, Weichen	Charité Universitätsmedizin Berlin
Löffler, Diana	Universität Würzburg
Lorenz, Romy	Klinische Navigation und Robotik Charité Universitätsmedizin Berlin, Campus Virchow Klinikum
Lotz, Peter	Continental Automotive GmbH
Luckner, Robert	Technische Universität Berlin, Institut für Luft- und Raumfahrt, Fachgebiet Flugmechanik, Flugregelung und Aeroelastizität
Luz, Maria	Technische Universität Berlin
Lylykangas, Jani	Research Group for Emotions, Sociality, and Computing, Tampere Unit Computer-Human Interaction, School of Information Sciences, University of Tampere, Tampere, Finland
Maier, Thomas	Universität Stuttgart
Mandiau, René	LAMIH-UMR CNRS 8201, University of Valenciennes, Le Mont Houy, France
Manske, Peer	Deutsches Zentrum für Luft-und Raumfahrt
Mantzke, Odette	Universität Bamberg
Manzey, Dietrich	Technische Universität Berlin

Marangé, Pascale	Centre de Recherche en Automatique de Nancy, CRAN, CNRS UMR 7039, University of Lorraine, Faculté des Sciences, France
Maugg, Franziska	Fraunhofer-Institut für Produktionstechnik und Automatisierung
Maus, Alexander	Automotive Safety Technologies, Wolfsburg
Mayer, Marcel P.	Lehrstuhl und Institut für Arbeitswissenschaft der RWTH Aachen
Meier, Sonja	Institut für Arbeitswissenschaft, RWTH Aachen
Minge, Michael	Technische Universität Berlin
Möhlenbrink, Christoph	Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)
Möller, Sebastian	Quality and Usability Lab, Telekom Innovation Laboratories, TU Berlin
Mühlstedt, Jens	Technische Universität Chemnitz
Müller, Simon	Technische Universität Berlin
Müller, Stephan	BMW Group, München
Müller-Plath, Gisela	Technische Universität Berlin
Nachtwei, Jens	Humboldt-Universität zu Berlin
Naujoks, Frederik	Universität Würzburg
Neukum, Alexandra	Interdisziplinäres Zentrum für Verkehrswissenschaften an der Universität Würzburg (IZVW)
Nielen, Alexander	Institute of Industrial Engineering and Ergonomics at RWTH Aachen
Nitsch, Verena	Universität der Bundeswehr München
Nurkka, Piia Eliisa	Tampere University of Technology, Finland
Oeltze, Katharina	Deutsches Zentrum für Luft- und Raumfahrt, Institut für Verkehrssystemtechnik
Oeser, Hendrik	Technische Universität Dresden, Professur für Verkehrspsychologie
Oja, Tuomas	Tampere University of Technology, Department of Pervasive Computing, Tampere, Finland
Ostheimer, Anne	TU Berlin, FG Mensch-Maschine-Systeme

Pacaux-Lemoine, Marie-Pierre	University of Valenciennes, France
Paschke, Jörg	Humboldt-Universität zu Berlin
Pétin, Jean-F.	Centre de Recherche en Automatique de Nancy, CRAN, CNRS UMR 7039, University of Lorraine, Faculté des Sciences, France
Pfeffer, Stefan	Universität Stuttgart
Philippot, Alexandre	CRéSTIC URCA, UFR Sciences Exactes et Naturelles, Moulin de la Housse, France
Polet, Philippe	University of Valenciennes, France
Popieul, Jean-Christophe	University of Valenciennes, France
Poulain, Thierry	University of Valenciennes, France
Qiu, Siqi	Compiegne University of Technology, France
Radüntz, Thea	Bundesanstalt für Arbeitsschutz und Arbeitsmedizin
Rachedi, Nedjemi	University of Valenciennes, France
Rajaonah, Bako	University of Valenciennes, France
Rantanen, Ville	Department of Automation Science and Engineering, Tampere University of Technology, Finland
Raudszus, Dominik	Institut für Kraftfahrzeuge, RWTH Aachen University
Reckin, Ronny	Technische Universität Berlin
Reich, Diana	Technische Universität Berlin, GRK Prometei
Reichmuth, Ralf	Technische Universität Berlin, GRK Prometei
Riedel, Laura	Technische Universität Berlin, Graduiertenkolleg prometei
Rinkenauer, Gerhard	IfADo - Leibniz Research Centre for Working Environment and Human Factors
Roßnagel, Alexander	Universität Kassel, Fachgebiet Öffentliches Recht
Roßner, Patrick	Technische Universität Chemnitz
Rötting, Matthias	Technische Universität Berlin
Ruff, Stefan	Technische Universität Berlin

Russwinkel, Nele	Technische Universität Berlin
Sachse, Katharina	Technische Universität Berlin
Salmimaa, Taru	Tampere University of Technology, Finland
Schindler, Julian	Deutsches Zentrum für Luft- und Raumfahrt
Schippert, Katharina	Hochschule der Medien (HdM), Stuttgart. Information Experience and Design Research Group
Schlag, Bernhard	Technische Universität Dresden
Schleicher, Robert	Telekom Innovation Laboratories
Schlick, Christopher M.	Lehrstuhl und Institut für Arbeitswissenschaft der RWTH Aachen
von Schlippe, Max	Technische Universität Berlin
Schmidt, Henning	Forschungsgruppe Rehabilitationsrobotik (IPK/TU Berlin), Abteilung Automatisierungstechnik, Fraunhofer Institut für Produktionsanlagen und Konstruktionstechnik (IPK), Berlin
Schmidt, Ludger	Universität Kassel
Schmuntzsch, Ulrike	Technische Universität Berlin
Schöning, Sebastian	Fraunhofer IPA
Schreiter, Karolin	Technische Universität Berlin, Institut für Luft- und Raumfahrt, Fachgebiet Flugmechanik, Flugregelung und Aeroelastizität
Schuldt, Michaela	Universität Kassel
Schwarze, Anke	Technische Universität Braunschweig
Silvennoinen, Johanna Maria	University of Jyväskylä, Finland
Simon, Philippe	UVHC, France
Skottke, Eva-Maria	HMKW Köln
Smieszek, Hardy	Technische Universität Berlin, GRK Prometei
Söllner, Matthias	Universität Kassel, Fachgebiet Wirtschaftsinformatik
Špakov, Oleg	Research Group for Emotions, Sociality, and Computing, Tampere Unit for Computer-Human Interaction, School of Information Sciences, University of Tampere, Tampere, Finland

Stark, Rainer	Fraunhofer IPK
Surakka, Veikko	Tampere Unit for Computer-Human Interaction, School of Information Sciences, University of Tampere, Finland
Temme, Gerald	Deutsches Zentrum für Luft- und Raumfahrt
Thüring, Manfred	Technische Universität Berlin
Tille, Ralph	Hochschule der Medien Stuttgart
Timpe, Klaus-Peter	Technische Universität Berlin
Totzke, Ingo	Würzburger Institut für Verkehrswissenschaften (WIVW GmbH)
Tuisku, Outi	University of Tampere, Finland
Tunca, Ercan	Universität Kassel, Fachgebiet Mensch-Maschine-Systemtechnik
Vääätäjä, Heli	Tampere University of Technology, Finland
Vanderhaegen, Frédéric	University of Valenciennes, France
Venesvirta, Hanna	University of Tampere, Finland
Venjakob, Antje	Technische Universität Berlin
Verho, Jarmo	Sensor Technology and Biomeasurements, Department of Automation Science and Engineering, Tampere University of Technology, Tampere, Finland
Vetek, Arkos	Nokia Research Center, Finland
Vick, Axel	TU Berlin, FG Industrielle Automatisierungstechnik
Vilimek, Roman	BMW Group, Anzeige- und Bedienkonzepte, München
Vogel, Marlene	Technische Universität Berlin
Voigt, Anja	Technische Universität Berlin, Institut für Luft- und Raumfahrt, Fachgebiet Flugmechanik, Flugregelung und Aeroelastizität
Volk, Madeline	Universität Würzburg
Wacker, Arno	Universität Kassel, Fachgebiet Angewandte Informationssicherheit
Wäfler, Toni	Fachhochschule Nordwestschweiz FHNW, Hochschule für Angewandte Psychologie, Olten

Wandke, Hartmut	Humboldt Universität zu Berlin
Warkentin, Jana	RWTH Aachen University, Aachen
Weber, Lennart	EADS Deutschland GmbH
Weller, Gert	Technische Universität Dresden
Werk, Sebastian	Technische Universität Berlin
Wetzel, Inge	EADS
Wienrich, Carolin	Technische Universität Berlin
Wieringa, Peter A.	Delft University of Technology, Faculty of Mechanical Engineering, BioMechanical Engineering, Netherlands
Wille, Matthias	Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA)
Wittenberg, Carsten	Hochschule Heilbronn
Wortelen, Bertram	OFFIS - Institute for Information Technology, Oldenburg
Zickerow, Matthias	Technische Universität Berlin
Zimmermann, Jasmin	Fachhochschule Nordwestschweiz (FHNW), Schweiz
Zlocki, Adrian	Forschungsgesellschaft Kraftfahrwesen mbH Aachen
Zoller, Ingo	Continental Automotive GmbH
Zwahlen, Pia	Schweizerische Bundesbahnen SBB, Schweiz